THE INEFFABILITY OF PERCEPTUAL EXPERIENCE: EXPLAINING THE CENTRAL INTUITION BEHIND JACKSON'S KNOWLEDGE ARGUMENT

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Emily Lynn Esch

January 2008

THE INEFFABILITY OF PERCEPTUAL EXPERIENCE: EXPLAINING THE CENTRAL INTUITION BEHIND JACKSON'S KNOWLEDGE ARGUMENT

Emily Lynn Esch, Ph. D.

Cornell University 2008

Many people share the intuition that in order to know what pineapple tastes like, one must have tasted pineapple. This is a special instance of the more general intuition that there are truths which cannot be understood unless one has had certain experiences. The influence of this intuition is widespread; it underlies debates about concept acquisition, the individuation of sense modalities, and our knowledge of the external world and of other minds. In this dissertation, I examine a famous argument in which this intuition plays a prominent role, Frank Jackson's knowledge argument. I articulate and defend this general intuition, and offer a diagnosis of its source: the reason these truths cannot be understood by someone who has not had the relevant experiences is that the content of visual experience is partly ineffable, i.e., the content of visual experience cannot be fully expressed linguistically. I distinguish between two types of ineffability, argue that only one of them would explain the central intuition, and isolate a feature of our visual experience that could explain why our experiences are necessarily ineffable. I argue that if the ineffability of perceptual experience is part of an explanation of the central intuition, then it must be interpreted as a claim about the phenomenal nature of experiences, rather than as a claim about the abilities of the subjects of experience. After close examination of three of the most prominent phenomenological features of perceptual experience – richness, fineness of grain, and determinacy – I argue that only determinacy is an essential property of

phenomenological content, and thus the right type of property to explain why experiences are necessarily ineffable.

BIOGRAPHICAL SKETCH

Emily Esch grew up in Joplin, Missouri. She graduated from Reed College in 1997 and began her graduate studies at Cornell in 1999. Emily and her husband, Gavin Henry, live with their two dogs and cat in Saint Cloud, Minnesota. Emily is currently an Assistant Professor of Philosophy at College of Saint Benedict/Saint John's University.

To Esmé, who was with me from beginning to end

ACKNOWLEDGMENTS

I would like to thank the following people for their support during the years I spent writing this dissertation. I begin with a thank you to the faculty and graduate students at Cornell, for both their intellectual and social companionship. I'd especially like to thank: Andrea Apostol, Eric Eben, Matti Eklund, Eric Gilbertson, Carl Ginet, Matt Haug, Benj Hellie, Eric Hiddleston, Brendan Jackson, Daniel Koltonski, Emily Muller, Peter Sutton, Zoltán Gendler Szabó, Brian Weatherson, Jessica Wilson, Pekka Väyrynen, and Aaron Zimmerman.

Thanks to all my friends at the Chanticleer and in particular to the bartender, Becky. My friends Kate Dunn and Becko Copenhaver deserve a special thanks for always being willing to offer encouragement and advice. I am extremely grateful to Anne Nester, who provided invaluable and extensive comments on countless drafts over the years. I'd also like to thank my family for their patience throughout this process, especially my incredible husband Gavin Henry, who has been a source of constant support for me over the last five years.

Finally I'd like to extend my thanks to my special committee, Sydney Shoemaker and Dick Boyd, and especially my advisor, Tamar Szabó Gendler. Without Tamar's encouragement and always sound advice I would not have completed this dissertation. I can't even begin to express all the ways I've learned from her over the past few years.

TABLE OF CONTENTS

Chapter One

Introduction

In a 1982 paper, "Epiphenomenal Qualia," Frank Jackson asks us to imagine the following situation:

> Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black-and-white room via a black-and-white television monitor. She specialises in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like 'red', 'blue', and so on…What will happen when Mary is released from her black-and-white room or given a colored television monitor? Will she *learn* anything or not? (1982, p. 42) [1]

Jackson claims that Mary will learn something when she leaves her room for the first time. His argument, which has come to be known as "the knowledge argument," can be stated more formally as follows:

> (KA1) Before leaving her room, Mary knows all the physical truths about color perception.
>
> (KA2) Before leaving her room, there is a truth about color perception Mary does not know.
>
> (KA3) Thus, there is a truth about color perception that is not a physical truth.

If physicalism is the thesis that all truths about the world are physical truths, then the knowledge argument shows physicalism to be false. For there is some truth about the world – namely, the truth about color perception that Mary does not know – which is not a physical truth. Jackson originally presented the knowledge argument as a

[1] See also Jackson (1986). Jackson has since changed his mind about the knowledge argument (1998, 2002). These three articles are reprinted in *There's Something About Mary* (eds. Ludlow, Nagasawa, and Stoljar). All page references are to this volume.

challenge to physicalism, and much of the discussion in the intervening quarter-century has focused on this purported implication.

My primary concern in this dissertation is with the substantive assumption about the nature of perceptual experience that underlies premise (KA2). This, which I will call the *central intuition* behind the knowledge argument, is the assumption that Mary must leave her room in order to learn certain truths about color perception. The central intuition has been denied, but most of the responses to the knowledge argument have tried to accommodate it. Few philosophers, however, have explained in any detail why it might be true. This dissertation is an attempt to give just such an explanation.

But before I begin my investigation, in Chapter Two, into what might lie behind the central intuition, I want to clear some ground. While there are not many who believe that the knowledge argument is successful, there are many different explanations of how the argument goes wrong.[2] Rather than try to cover all the material that's been written on the knowledge argument, in this chapter I focus on three broad topics which have been influenced by the knowledge argument: the nature of physicalist explanation, the explanatory gap, and phenomenal concepts.[3] As this first chapter shows, the knowledge argument continues to generate interest because it involves so many of the perplexing problems of contemporary philosophy of mind. I, however, have tried to focus my attention on the issues surrounding the reasons why so many have found it plausible that Mary must have the experience of seeing red

[2] See Van Gulick (2004), Chalmers (2004), and Stoljar and Nagasawa's Introduction (2004) in *TSAM* for surveys of possible responses to the knowledge argument.

[3] The literature on the knowledge argument is wide-ranging and multiplying and I can't do justice to all of it. As I was finishing this dissertation a new volume on issues related to the knowledge argument was published, *Phenomenal Concepts and Phenomenal Knowledge* (2007) eds. Torin Alter and Sven Walters. I regret that I was not able to incorporate more the issues which come up in this collection.

before she can know what seeing red is like. This chapter is designed to show what, if any, bearing the central intuition has on these three topics.

All of the physicalist responses that I will consider in this chapter grant that the thought experiment Jackson proposes is coherent, and so they accept some form of the intuition behind the second premise. These physicalists concede that Mary will gain something when she leaves her room, though descriptions of what she gains vary. Given this concession, the physicalists' task is to explain why the possibility of Mary gaining new knowledge doesn't pose a problem.

Completing this task has motivated philosophers to refine their conception of physicalism. I show that different conceptions of physicalist theory have very different consequences for how one chooses to respond to the knowledge argument. I argue that a convincing response to the knowledge argument will depend on resolving the disagreement between a priori and a posteriori physicalism.[4]

The knowledge argument is also partly responsible for some philosophers' decision to abandon the belief that an adequate physicalist explanation of the mental requires the possibility of an a priori reduction of the mental to the physical. There is an "explanatory gap" that appears when we try to explain how phenomenal states can be identical to physical states, and many philosophers take the knowledge argument as a vivid illustration of this gap.

Finally, the knowledge argument has moved discussion forward about what are now usually called "phenomenal concepts." These are the concepts we use when introspecting our experiences. The special nature of these concepts explains why Mary cannot know what it's like to see red while still in her black and white room—she lacks the conceptual resources to understand certain claims. Mary can't acquire

[4] The complicated issues involved in the debate between the two physicalisms strike to the core of how we should understand philosophical methodology. This is one reason why I have not tried to adjudicate between the two positions in this dissertation.

these concepts until she has had the relevant experience, so despite her complete scientific knowledge she can't know what it's like to see red.

As we'll see in more detail, the three topics are frequently interconnected, although it's not always easy to see the precise nature of the connections. How one understands the commitments of physicalism will affect how seriously one takes the threat of the explanatory gap. And many philosophers have argued that it is the special nature of phenomenal concepts that explains the appearance of a gap between the physical and the mental.

1.1 Physicalism

In this section, I want to look at two arguments, by Terence Horgan and Daniel Stoljar, which propose that there are ways to formulate physicalist theory that allow it to evade the conclusion of the knowledge argument.[5] Both Horgan and Stoljar accept that Mary gains something when she leaves her room and sees red for the first time; in other words, they both agree that she would not be able to deduce knowledge of what seeing red will be like from the scientific knowledge she has. And these two arguments share a similar structure: first, they draw a distinction between two interpretations of "physical," and second, they claim that the knowledge argument's apparent validity rests on a failure to disambiguate between the two notions.

But the implications of the two responses are very different. Horgan thinks the problem posed by the knowledge argument is primarily epistemological: Mary can't know what it's like to see red until she sees it, because this kind of knowledge requires being in a special epistemic relationship to the experience of red. Stoljar's contribution is primarily metaphysical; he describes a type of physicalism which

[5] See Horgan, Terence (1984). "Jackson on Physical Information and Qualia," *The Philosophical Quarterly* 34: 147-152. Reprinted in *TSAM*. Stoljar, Daniel (2001) "The Conceivability Argument and Two Conceptions of the Physical" *Philosophy and Phenomenological Research* 62: 253-281. An abbreviated version of this article is reprinted in *TSAM*. All page references are to this volume.

allows for more properties than those ascribed by scientific theory. These categorical properties, combined with the dispositional properties Mary already knows about, form the supervenience base of the phenomenal properties.

1.1.1 Horgan: The Epistemic Response

Horgan (1984) begins by distinguishing between truths that are *explicitly physical* and truths that are merely *ontologically physical*.[6] He defines *explicitly physical truths* as truths that are either expressed by the sentences of an adequate scientific theory of a physical process or follow from such a theory. And he defines *ontologically physical truths* as truths in which all of the entities referred to, or quantified over, are physical (304). Since, by hypothesis, Mary knows all the scientific theories before leaving her room, Mary knows all the explicitly physical truths. But it doesn't follow from this, according to Horgan, that she knows all the ontologically physical truths. Some truths about physical entities will not be expressed in explicitly physical language; one of these will be the truth about what it's like to see a red thing. This is no problem for physicalism since before and after Mary sees red, she refers to the same physical entity (this might be a functional state, a neurological state, etc., depending on the theory).

All that the physicalist should commit herself to, says Horgan, is the claim that the world is composed of only physical things; a physicalist does not need to be committed to the view that all sentences can be translated into the explicitly physical language of science. The knowledge argument trades on an equivocation between the explicit and ontological notions of "physical truths." To see this equivocation, let's

[6] Horgan distinguishes between kinds of *information*, not kinds of truths. But since I've formulated the knowledge argument in terms of truths, for consistency I've reformulated Horgan's objection. I don't think that this has any bearing on the strength of the objection. Both Horgan (1984) and Stoljar (2001) are reprinted in *TSAM*, and all page references are to this volume.

first understand “the physical truths” that Mary knows before leaving her room to mean the explicitly physical truths. The reformulated argument is this:

(1H) Before leaving her room, Mary knows all the explicitly physical truths about color perception.

(2) Before leaving her room, there is a truth about color perception she does not know.

From these premises, (3) does not follow:

(3)* Thus, there is a truth about color perception that is not a physical truth.

It would follow only with the addition of a premise such as the following:

(2.5H) All physical truths about color perception are explicitly physical truths.

As it stands, the argument establishes only:

(3H) Thus, there is a truth about color perception that is not an explicitly physical truth.

Horgan denies that the physicalist is committed to (2.5H); he thinks there are ontologically physical truths that cannot be expressed in explicitly physical language. So, despite the fact that she knows all the explicitly physical truths, Mary still doesn’t know what it’s like to see red.

When the knowledge argument is stated in terms of ontologically physical truths, the first premise is:

(1H’) Before leaving her room, Mary knows all the ontologically physical truths about color perception.

Horgan argues that physicalists should deny (1H’) (305-6). Mary cannot know all the ontologically physical truths before leaving her room, and physicalists are not committed to the view that she does. There will, according to this line of thinking, be

certain truths that cannot be expressed by explicitly physical language.[7] One of these ontologically physical truths will express what the experience of seeing red is like. Mary can't know this truth until she leaves her room, and so the first premise should be denied.

If Horgan is correct about the equivocation, then the knowledge argument is invalid, and clearly not a threat to physicalism. Yet there remains a nagging question: can the physicalist offer a positive account of Mary's new knowledge? Physicalism is supposed to be a complete theory, and as such it needs to account for the kind of knowledge Mary gains when she leaves her room.[8]

Horgan thinks that the physicalist can account for this new knowledge by appeal to the special epistemic access Mary now has to the experience of seeing red. Before leaving her room, Mary had a complete scientific description of the experience, which is one way we have of conceiving the experience. When Mary leaves her room, however, she gains another way of conceiving of the experience, one that is made possible by her new introspective access to the experience. What changes in Mary before and after she leaves her room is her epistemic relation to the experience of seeing red. The experience itself is the same physical state.

1.1.2 Stoljar: The Metaphysical Response

Daniel Stoljar (2001) distinguishes between two kinds of physicalism, the *theory-based conception* and the *object-based conception*, which correspond to distinct views of physical properties. On a theory-based conception of physicalism,

---

[7] These truths cannot be expressed linguistically, period. For Mary doesn't just know scientific descriptions of the phenomenon, she also has first hand reports from ordinary people. These reports will not be expressed in explicitly physical language, but they also won't help her to learn what it's like to see red.

[8] Jackson has repeatedly stressed this point. See his (1986), (1998), (2004). But Jackson's notion of completeness is controversial; for him completeness requires the possibility of an a priori reduction of the phenomenal concepts to the physical concepts.

physical properties are those properties that are either mentioned by physical theories, or supervene on properties that are mentioned by the physical theories. On an object-based conception of physicalism a physical property is a property that is either required by a "complete account of the intrinsic nature of paradigmatic physical objects" (e.g. rocks and flowers) or supervenes on these properties (312).

Stoljar argues that combined with two additional claims, these two conceptions of physicalism pick out different classes of properties. The first claim concerns the nature of scientific theories. Stoljar recognizes a view of scientific theories in which physical theories refer only to dispositional properties, properties like having a certain mass or being fragile—I'll call this claim about scientific theories *dispositionalism*.[9] The second claim is that dispositional properties require categorical grounds: for example, the fragility of the window, a dispositional property, is grounded in the molecular structure of the glass it's made of, a categorical property that provides grounds for the disposition. I'll call the claim that dispositions require categorical grounds *categoricity*.[10]

Stoljar argues for the distinctness of the two conceptions of physicalism in the following way. Let's assume that dispositionalism and categoricity are true. We can now see that the conceptions of physicalism yield different classes of properties. Suppose that there is an object with a dispositional property G, which counts as a physical property under both the object-based and the theory-based conceptions. By applying categoricity, it is known that G has a categorical base property F. By applying dispositionalism, it follows that on the theory-based conception of

[9] Stoljar quotes Simon Blackburn (1990) as a proponent of this position: "science finds only dispositions all the way down." See also Armstrong (1968).

[10] While they might first appear to be in conflict, these claims are consistent. Dispositionalism is a claim about the nature of scientific explanation; categoricity is a claim about the metaphysics of properties. Of course, if you think that only the properties which are mentioned by physical theories exist, then you will think the second claim is false.

physicalism, F does not count as a physical property. On the object-based conception of physicalism, however, it is at least possible that F counts as a physical property. Whether or not it does will depend upon whether F is needed in a complete account of the intrinsic nature of paradigmatic physical objects. Let's assume that it is.[11] Then there is a property which counts as physical according to the object-based conception, but not according to the theory- based conception. Or, to use Stoljar's phrase, the object-based conception of physicalism "tells us about" a property that the theory-based conception of physicalism does not.[12]

Stoljar argues that distinguishing between the two conceptions of physicalism provides a response to the knowledge argument (317-319). On the one hand, if we assume the theory-based conception of physicalism, then the conclusion of the knowledge argument doesn't follow. Here is the new argument:

(1S) Before leaving her room, Mary knows all the truths about theory-based properties involved in color perception.

(2) Before leaving her room, there is a truth about color perception she does not know.

(3) *Thus, there is a truth about color perception that is not a physical truth.

As above, the argument isn't valid without the addition of the following premise:

(2.5S) All the truths about color perception are theory-based truths.

As it stands, the argument only establishes:

[11] Stoljar (2001) doesn't explain what a complete account of the intrinsic nature of paradigmatic physical objects would look like, or how we would know it is complete. For my purposes, it's sufficient to leave this notion vague. See his (2005) for more discussion on this topic.

[12] Stoljar claims that the failure of a physical theory to tell us about categorical properties is not a failure of reference. Instead he argues that a physical theory fails to tell us about categorical properties if and only if it is possible that there are two worlds that are identical in the distribution of their dispositional properties and are epistemically and qualitatively identical from the point of view of the theory, but have different categorical grounds. See p. 315.

(3S) Thus, there is a truth about color perception that is not a truth about theory-based properties.

Stoljar argues that (2.5S) is false. The category of object-based physical properties is larger than the category of theory-based properties. And (3S) doesn't worry him, because the object based properties are also required for complete knowledge of color perception. Mary's complete stock of scientific truths will not provide her with the information about categorical properties necessary for a complete characterization of the intrinsic nature of physical objects.

On the other hand, if we assume the object-based conception of physicalism, then, Stoljar argues, there is no reason for the physicalist to accept (1S'):

(1S') Before leaving her room, Mary knows all the truths about object-based properties involved in color perception.

According to dispositionalism, physical theory won't provide the truths about all the object-based properties (i.e. it won't mention the categorical properties). So she doesn't know all the object-based truths before leaving her room; scientific descriptions won't provide her with any truths about object-based properties.

Of course, for this response to be effective in blocking the conclusion of the knowledge argument, we must accept not only that there is a distinction between the two conceptions of physicalism, but also that the object-based conception is the correct one. For only the object-based conception allows us to deny that Mary knew all the physical truths before she left her room. If the theory based conception of physicalism were correct, then Mary, given her complete knowledge of the scientific theories of color perception, would be in a position to figure out what seeing red is like. Since science isn't in the business of providing information about categorical properties, no matter how much scientific knowledge she has she will never have enough to provide her with knowledge of what seeing red is like.

Let's grant for the moment that the object-based conception of physicalism is the one we should adopt. So one of the truths Mary will be missing despite her mastery of the scientific truths is a truth about an object-based property, a categorical ground for one of the dispositional properties. It's important to recognize that this truth about the categorical property is not what Mary learns when she leaves her room and sees red. Because of his desire to defend a form of physicalism, Stoljar doesn't think that the categorical property is identical to the phenomenal property.[13]

How do these properties help the physicalist respond to the knowledge argument? Stoljar explains:

> [P]hysical theory does not tell us everything about the physical world: it is selective and only tells us about dispositional t-physical properties when in addition there are categorical o-physical properties. Of course the categorical o-physical properties are not themselves qualia. But in combination—perhaps also in combination with the t-physicals—they may constitute qualia. (2001, 325)

If the categorical properties combine with the dispositional properties to form the supervenience base for phenomenal properties, then this explains why Mary could not work out, from inside the black and white room, what red things look like. For all she knows about in her room are the dispositional properties. Stoljar's response gives us a reason for thinking that Mary will be surprised when she sees red for the first time. Despite her complete knowledge of the scientific facts, she is missing pertinent information about the world.

Both Horgan and Stoljar point out that the commitments of physicalism are weaker than the knowledge argument seems to suppose. Horgan distinguishes between sentences that use explicitly physical language and those sentences that use only ontologically physical language, and argues that physicalism is not committed to

[13] Identifying the categorical properties with the phenomenal properties would result in panpsychism—the view that the categorical properties which ground the dispositional properties are always phenomenal properties—which is not compatible with physicalism.

the reductionist claim that the ontologically physical sentences reduce to the explicitly physical sentences, but is committed to only the weaker claim that all the referents be physical entities. Stoljar distinguishes between two types of physical properties and argues that physicalism is not committed to the view that before she leaves her room Mary knows all the truths about every type of physical property, but only the weaker claim that she knows all the truths about theory-based properties. Without knowledge of all the properties, she will be unable to figure out what seeing red is like.

In both these cases, once the notions are properly disambiguated, on one reading the knowledge argument's conclusion fails to follow. And on the other reading, when we understand 'physical truths' as referring to ontological or object-based truths, both Horgan and Stoljar argue that the physicalist should deny the premise that Mary knew all the physical truths. According to Horgan and Stoljar, there is no reason to think that Mary can know all the ontologically and object-based physical truths before she leaves her room. There are going to be certain truths that Mary doesn't know until she see colors for the first time.

Though the structures of the arguments are similar, Horgan and Stoljar see the challenge posed by the knowledge argument very differently. While Stoljar's proposal explains why Mary couldn't deduce the relevant facts about what seeing red is like, it doesn't tell us why she couldn't have been told about them. Stoljar focuses on Mary's special epistemic powers in a way that Horgan does not. Unlike Horgan, Stoljar is an a priori physicalist; that is, he accepts that truths about mental facts can be deduced a priori from truths about the physical facts. Given this commitment, Stoljar needs to explain why Mary is not in a position to do so, given her complete scientific

knowledge. His explanation, as we have seen, is that despite all of her scientific knowledge she doesn't know all the relevant facts.[14]

Horgan doesn't focus on Mary's scientific prowess, because he doesn't think that physicalism is committed to such an a priori reduction. Horgan's explanation of why Mary can't know all the facts in her room would apply equally well to why any person stuck in a black and white room would be unable to know what it's like to see red. None of us, until we have had the experience of seeing red or some similar color are able to know what the experience is like. According to Horgan, what we are missing is the special perspective which comes from having introspective access to the property.

Whether or not the knowledge argument succeeds as an argument against physicalism will ultimately depend on the nature of the supervenience relation that holds between the mental and the physical. Horgan think that our knowledge of the relation is empirical, while Stoljar thinks that it is a priori. And this dispute, between the a posteriori and the a priori physicalists, is fundamentally a dispute about how to close the explanatory gap.

1.2 The Explanatory Gap and Phenomenal Concepts

1.2.1. The Gap

The thought experiment behind the knowledge argument is often mentioned as an illustration of the explanatory gap. Over the last twenty years or so, the nature of the explanatory gap has been articulated and its persistence defended most vigorously

[14] I also want to point out that Stoljar's proposal offers no positive proposal for what Mary learns when she leaves her room. For knowing about the categorical properties is unnecessary for knowing what it is like to see red. All of us who can see red know what it's like to see red without knowing anything about categorical properties.

by Joseph Levine.[15] But the gap is not new. It is related to much earlier versions of the mind-body problem, as illustrated in this passage from Leibniz:

> Moreover, we must confess that the *perception*, and what depends on it, is *inexplicable in terms of mechanical reasons*, that is, through shapes and motions. If we imagine that there is a machine whose structure makes it think, sense, and have perceptions, we could conceive it enlarged, keeping the same proportions, so that we could enter it, as one enters into a mill. Assuming that, when inspecting its interior, we will only find parts that push one another, and we will never find anything to explain a perception. (1714/1991, p. 70)

Leibniz's machine pulls at our metaphysical intuitions: how can the stuff of our conscious lives—the smell of coffee, the feel of a summer breeze, the taste of a madeleine—be identical to our brain, however complicated its structure might be? But Leibniz uses these metaphysical intuitions to make an epistemological point: we have no conception of how conscious experiences could be explained by mechanical parts.

Like Leibniz's, Levine's explanatory gap is, as its name implies, a gap in our understanding of how the phenomenal character of our experiences could be identical to physical properties. The gap is frequently explicated by an analogy to our scientific understanding of the identity that holds between the scientific understanding of H_2O and the ordinary conception of water, in which there is no such explanatory gap. From a complete understanding of the scientific facts, it is argued, one can understand why water has its macroscopic properties, e.g. being colorless, odorless, liquid at room temperature, etc. By contrast, the relationship that holds between phenomenal states and brain states seems completely arbitrary. Why should this bit of neural tissue be identified with the experience of seeing red rather than blue or, more generally, with any conscious experience at all?

[15] See Levine (1983), (2001), and (2007).

It's obvious why philosophers would associate Mary's predicament with the explanatory gap. In Jackson's thought experiment, Mary is assumed to have complete knowledge of all the physical facts relevant to understanding human color perception, yet (almost) everyone believes that she fails to know everything there is to know about colors. Mary's ignorance is often attributed to her inability to derive (or deduce) knowledge of what it's like to see colors from the physical information at her disposal. Again, it's helpful to compare this to the identity which holds between water and H_2O. Imagine a parallel thought experiment. Martin is a brilliant scientist specializing in hydrology who is kept in a room and has never seen water. Like Mary, Martin knows all the physical facts relevant for a complete understanding of water behavior. Unlike Mary, let's assume that Martin doesn't know anything about the ordinary concept of water. What happens when Martin leaves the room for the first time and sees water? Will he be surprised to find that it is an odorless, colorless, and boils at 212 degrees? It seems that none of these properties of water will be a surprise to Martin. And so it seems that there is an important difference between our understanding of the identity that holds between H_2O and water and the identity that holds between phenomenal states and brain states.[16]

What type of physicalism you advocate is relevant to how seriously you take both the knowledge argument and the explanatory gap. The a priori physicalist is likely to be very much bothered by these sorts of cases.[17] But the a posteriori physicalists have tended to believe they can respond to both the knowledge argument and the explanatory gap. As Levine points out, the a posteriori physicalist generally

[16] There is, of course, much more to be said on behalf of the physicalist. A full defense of the explanatory gap is far beyond the scope of this dissertation, but I think that Levine continues to make a convincing case that the gap persists. My purpose here is merely to acknowledge the connections between the explanatory gap and the knowledge argument.

[17] For example, David Lewis's response to the knowledge argument, which is discussed in detail in Chapter Three, is driven by his commitment to a priori physicalism. It also lies behind Jackson's struggles with the knowledge argument.

responds to the challenge the explanatory gap poses in two ways (2007, 146-7). First, she denies that sentences containing the word "water" can be derived from the scientific descriptions; the vocabulary of science and the vocabulary of ordinary language are too different.[18] In other words, there is a gap in the ordinary case, too. Second, the a posteriori physicalist points out that it's in the nature of identity statements to not need explanation: when you have an identity statement, you've hit the end of the road. Asking how it's possible that the phenomenal state of experiencing red is a brain state is like asking how it's possible that water is identical to H_2O.

I agree with Levine that this is not enough. There still seems to be a difference in how we understand the identity between H_2O and water and the identity between phenomenal character and neurological properties. As Levine asks: "Why, in this case, does it seem so bizarre to consider what is picked out by the one vocabulary to be the very same thing as what is picked out by the other vocabulary, when no such bizarreness attends other theoretical reductions?" (2007, 148). Many philosophers have claimed that the bizarreness of the identity is due to the special nature of phenomenal concepts, the topic to which we now turn.

1.2.2 The Phenomenal Concept Strategy

The most popular physicalist response to the knowledge argument today also promises to explain away the problem of the explanatory gap.[19] Problems stemming

[18] Cf. Horgan's distinction between explicit and ontological physicalism discussed above.

[19] I offer only a sketch of the basic strategy, since there are many different ways to understand the nature of phenomenal concepts. In addition to the usual disagreements about how to understand the nature of concepts generally, there is debate over whether phenomenal should be understood as demonstrative, indexical, recognitional or functional concepts. See Horgan (1984), Perry (2001), Lycan (1996), Loar (1990), Tye (2002). As we've seen, Horgan thinks that Mary is missing a demonstrative: she can't think to herself "so *this* is what seeing red is like" while introspecting an experience of a red thing. Perry and Lycan endorse an indexical account; I discuss Lycan's view in detail in Chapter Four. Loar argues that phenomenal concepts are recognitional concepts of some kind and Tye argues that they are functional concepts.

from the knowledge argument and the explanatory gap will be resolved as soon as we recognize the radically special nature of phenomenal concepts. Phenomenal concepts are those concepts we use to pick out what it's like to be in a mental state, for example, to pick out what it is like to undergo the visual experience most humans have while seeing a red apple or to demonstrate that seeing red is like this while introspecting a red experience.[20] Stuck in her black and white room, Mary lacks all phenomenal concepts of color experiences. Thus, she can't know all the truth about colors.

The proponent of the phenomenal concept strategy relies on what I called in the beginning of the chapter, *the central intuition.* This is the intuition that Mary cannot have knowledge of what it is like to see red until she has the experience of seeing something red. Phenomenal concepts are special in that their possession requires having the relevant experience.[21] According to a proponent of this strategy, the problem facing Mary isn't so much that she can't *know* certain truths about color; she can't even entertain the right sort of thought, because she lacks the necessary concepts.

Those who believe phenomenal concepts offer the physicalist an explanation of the knowledge argument are a posteriori physicalists. The innocuous nature of Mary's new knowledge is explained by comparing the identity statement she comes to know to other scientific identity statements. According to the a posteriori physicalist, Mary's new knowledge, like her old knowledge, is about physical properties. Both before and after leaving her room, Mary's thoughts are about the phenomenal state, which is realized by a physical state of the brain. When pre-release Mary wonders

[20] This is awkward, I know. It is difficult to be precise about what these phenomenal concepts pick out. (Hence, the title and thesis of the dissertation.)
[21] They might be special in other senses, too, though it's hard to find a straightforward, noncircular account of what makes these concepts special in any other sense.

what it is like for ordinary people to see red, she is wondering what it is like to be in a certain phenomenal state R, and when she is herself seeing red she is thinking about what it is like to be in state R; in both cases her thoughts are about R.[22] What changes when Mary leaves her room is that she acquires concepts she couldn't possess while in the black and white room. These new phenomenal concepts allow her to entertain the appropriate knowledge claims.

I now turn to the question of how the central intuition to connected to a posteriori physicalism. The central intuition is obviously compatible with a posteriori physicalism, but is there a stronger, explanatory relation which holds between them? For example, does the central intuition explain why Mary can't deduce what it's like to see red from all the physical truths? A thought experiment proposed by Nida-Rumelin shows that this is not the case.[23] Imagine that one day Mary's black and white room is completely re-fitted in artificial colors. The walls, the curtains, the dishes, etc, are all replaced with variously colored items, but nothing in the room reveals an object's natural color. (Bananas do not appear to be yellow; Mary's own body remains black and white.) While Mary is free to enjoy her new surroundings, she is not told any of the names of the colors of the new objects in her room. After living in the colored room for a week, Mary is shown four color samples: red, blue, green, and yellow and asked to pick out the color which most resembles the color of the sky, the color of a ripe tomato, etc. It seems clear to many that Mary wouldn't know.[24]

---

[22] If the referent, in this case R, were the only element involved in individuating thought contents, then the physicalist would be stuck arguing that Mary does not gain any new information when she sees red. But there are good reasons for thinking that something else is involved in the case of intentional mental states like belief.

[23] Nida-Rümelin (1995). See also Stoljar (2005) and Levine (2007) for discussions of versions of this new thought experiment.

[24] I discuss those who believe Mary would know in Chapter Three.

Let's assume that Mary isn't able to correctly apply her concept PHENOMENAL RED to the ripe tomato; let's assume that she thinks that the blue chip accurately picks out the color of ripe tomatoes. Then the thought experiment shows that the experience thesis does not imply or explain in any way a posteriori physicalism.[25] In the new thought experiment, Mary has acquired the phenomenal concepts; she knows what it's like to see red, even if she doesn't know what other people call this experience. But possession of the phenomenal concepts alone does not allow her to derive the phenomenal knowledge she's missing. She still doesn't know what a ripe tomato or a clear sky looks like to normal perceivers.

It seems that the knowledge argument could be run using the new thought experiment to support premise (KA2). In addition to all her old knowledge Mary now has a bunch of phenomenal concepts. But, the intuition is that Mary will not be able to appropriately apply her phenomenal concepts and hence there will still be a truth about colors that Mary doesn't know, for example that a ripe tomato looks like the red color sample.[26]

Conclusion

In this chapter, I've introduced the knowledge argument and my primary interest in understanding why so many accept the second premise: (KA2) Before leaving her room, there is a truth about color perception Mary does not know. I've said that the reason behind this acceptance is the intuition that Mary must have experiences of colors before she can possess certain beliefs about them. This intuition has been around for a long time, and in the next chapter I explore some of its recent history.

[25] Stoljar (2005, p. 487) makes this argument, and Levine (2007) makes a similar point.
[26] I discuss the distinction between possessing a concept and applying a concept in more detail in Chapter Three.

I also discussed three topics which have come to be closely associated with the knowledge argument, the nature of physicalist explanation, the explanatory gap, and phenomenal concepts. I discussed these three issues for two reasons. First, I think the knowledge argument deserves credit for encouraging philosophers to think seriously about some of the most difficult problems in philosophy of mind. Secondly, I wanted to point out that, interesting as these issues are, they are not directly connected to the central intuition. In the chapters which follow, I try to investigate the central intuition on its own, without worrying about its implications for physicalism or phenomenal concepts.

Chapter Two
The History Behind the Central Intuition

Introduction

Most people believe that there are certain things one can't know until having the relevant experience. A man blind from birth, for example, cannot know what red looks like. A woman born deaf cannot know what the surf crashing sounds like. A boy who has never tasted a pineapple cannot know what pineapple tastes like. The intuition that having the relevant experiences is necessary for acquiring knowledge is most forcefully pressed by the canonical modern empiricists, John Locke, George Berkeley, and David Hume. While I don't endorse the original broad scope of the empiricist claim, I think the empiricists were correct about a certain class of these truths.

In this chapter, I examine the historical roots of empiricist beliefs about the nature of perceptual experience and the imagination, and how these beliefs support the claim that all knowledge is derived from sense experience. I begin chronologically, with John Locke's statement of what I'll call the *empiricist principle*:

> Whence has [the mind] all the *materials* of reason and knowledge? To this I answer, in one word, from EXPERIENCE. In that all our knowledge is founded; and from that it ultimately derives itself. Our observation employed either, about external sensible objects, or about the internal operations of our minds perceived and reflected on by ourselves, is that which supplies our understandings with all the *materials* of thinking. These two are the fountains of knowledge, from whence all the ideas we have, or can naturally have, do spring. (Bk II. Ch I. 2)

There are two points I want to make about this passage. First, there is the scope of Locke's claim: *all* knowledge is founded in experience. Second, as this passage makes clear, the empiricist principle is fundamentally a claim about the source of all

our ideas: all *materials* of knowledge come from experience. The first point is epistemological—our knowledge is founded, i.e. justified, by experience. The second point is semantic—the ideas which compose our beliefs come from experience.[1] The two points can be connected in the following way: in order to have knowledge that P, we must possess the ideas which compose P.

Locke's empiricist principle can be stated more succinctly as:

> (EP) All the materials of human knowledge are derived from experience, and all human knowledge is founded in experience.

Both Berkeley and Hume accept this basic statement of the empiricist principle, and each refine it in different ways. My interest lies primarily in the semantic claim. I share with these empiricists an interest in the source of our ideas and defend the view that in order to possess a certain class of ideas, we must first have the relevant experience. This dissertation is an attempt to understand the reasons why the empiricist principle might be true for certain kinds of beliefs.[2]

This chapter is structured chronologically around three thought experiments, each which shed light on a different aspect of the empiricist principle. The chapter begins with a discussion of Locke and his arguments in favor of the empiricist principle. These arguments are divided into two sorts: arguments against the nativist view that was dominant at the time Locke was writing and positive arguments which directly support the empiricist principle. Interestingly, one of the thought experiments Locke offers in support of his empiricism parallels the thought experiment behind the

[1] To contemporary readers, it often seems that the moderns were not careful of the distinction between the source of a belief P and justification for belief P. There does seem to be a tendency to conflate epistemic and semantic claims.

[2] As I mentioned, I do not endorse the broad scope of the empiricist principle. One of the issues discussed in this chapter is why the empiricists thought the empiricist principle applied to all our ideas. It turns out that many of their arguments center on the types of ideas (of colors, sounds, tastes and other sensory qualities) that are candidates for my own proposal offered in Chapter Four. In other words, their arguments are most convincing about a restricted class of ideas; the same set of ideas which are relevant to my own interests.

knowledge argument. While I think that Locke's arguments against the nativist are successful, his arguments in favor of the empiricist principle are too quick, relying primarily on introspective reports of our own mental operations and abilities.

In the second section, I turn to a thought experiment which generated great interest during the modern period, Molyneux's Question. William Molyneux, author of a treatise on optics, wrote to Locke and asked the following question: would a man, blind from birth who suddenly regained his sight, be able to identify, from vision alone, a cube and a sphere? Locke's own response to Molyneux is brief, but in *A New Theory of Vision*, George Berkeley lays out a view of perceptual experience that allows for a sophisticated treatment of the question. Berkeley realizes that a negative response requires the empiricist to adopt some fairly radical views on the natures of visual and tactile experiences and the relationship that holds between the two. My reasons for discussing Molyneux's Question in such depth is twofold: first, the differences between Locke's and Berkeley's responses provides a deeper understanding of nature of perceptual experience, and second, the situation described by Molyneux bears a certain affinity to Mary's situation as described by Jackson. In fact, as I discuss in Chapter Four, some philosophers have implicitly assumed that Mary's situation is merely a special instance of Molyneux's Question.

The final section of the paper is devoted to Hume's famous thought experiment concerning the missing shade of blue. After a discussion of the faculty of imagination, I compare Hume's contribution to the empiricist claim, the principle that all simple ideas are faint copies of simple impressions. Hume claims, contrary to his adherence to the copy principle, that a man facing a spectrum of blue shades from light to dark with a gap in the middle of the spectrum will be able to imagine an idea of the shade, even though he has never encountered that shade before. The missing shade of blue is worth discussing in detail since, first, it is perhaps the most famous counterexample to

the empiricist claim, and two, because the imagination plays an important role in how we understand the nature of phenomenal concepts, those concepts which allow us to form beliefs about what our perceptual experiences are like.

2.1 The Empiricist Principle

Locke begins an *Essay concerning Human Understanding* by arguing against the popular nativist view that humans are born possessing certain ideas and knowing certain principles. The nativists, who at the time Locke wrote the *Essay* were most closely associated with Descartes and his followers, believed that we are born with minds that have been supplied, by God or Nature, with ideas and principles. The scope of the nativist claim varies, but it typically includes the ideas of substance, God, identity and mathematical and logical principles.[3]

It's natural to suppose that Locke's arguments are a consequence of his empiricism. However, a close look at the arguments of Book I shows that the arguments against innate ideas and principles do not depend on a prior acceptance of the empirical claim that all our ideas originate in experience. Rather, Locke sets up a dilemma for the nativist and attempts to prove that the doctrine of nativism is either false or trivially true.[4] The success of this dilemma depends, of course, on whether Locke has fairly rendered the account of his opponents. I defend Locke against the charge of attacking a straw man at the end of this section.

[3] At one time, Descartes was willing to include even our ideas of sensation as innate. More on this below.
[4] See Atherton (1983), Jolley (1999) for discussions of the dilemma. My discussion of Locke is especially indebted to Atherton's interpretation.

2.1.1 The Nativist's Dilemma

One of the striking features of Book I is that Locke doesn't much discuss innate ideas until the brief third chapter.[5] Instead, the bulk of his arguments are directed towards innate principles, e.g. logical propositions like "That it is impossible for the same thing to be and not to be" and "Whatever is, is." One reason for this emphasis on propositions over ideas is that Locke is especially concerned with responding to the popular argument that there are certain maxims, e.g. those just mentioned, which command universal assent. One assents to propositions, and not to ideas.

Secondly, Locke assumes that innate propositions are composed of innate ideas: "no proposition can be innate unless the *ideas* about which it is are innate" (Bk I, Ch 2, 17). In order to know that it is impossible for the same thing to be and not to be, one must have the ideas of things, existence, etc. It's possible, of course, that there are no innate principles, but there are innate ideas. In other words, it's possible that humans are born with the basic components of the principles (the ideas), but require experiences to learn how to put the ideas in propositional form. This is not a possibility Locke discusses, but the possibility shows that arguments against innate principles don't automatically carry over to innate ideas.[6]

---

[5] Locke's use of the word "idea" is notoriously problematic; he uses it in a variety of different ways. See Ryle (1933), Chappell (1994) for examples of the many ways Locke uses the word. But it is clear that one use of the term is as the object of our thoughts, and it is this use that matters for my purposes.

[6] Given the empiricist principle, though, it seems likely that behind Locke's arguments against innate principles is the belief that these principles can't be formed until the ideas of which they are composed are acquired through experience. In what follows, the discussion moves back and forth between innate principles and innate ideas. I chose to present the arguments in this potentially confusing way for two reasons. The first reason is for historical accuracy; it would have been misleading to reformulate Locke's arguments against innate principles in terms of innate ideas. Second, I wanted to present the various accounts of innateness in the most compelling way. Some of the theories work better as theories of innate ideas and some work better as theories of innate principles. I am not, for the reason given above, claiming that ideas and principles are interchangeable in this context.

Locke's arguments for the first horn of the dilemma – that it is empirically false that we have innate principles – are straightforward. He argues, for example, that if there are innate, universal principles as the nativists have claimed, then these principles should be present in children. But, Locke claims, it is obvious that children don't assent to maxims like, "Whatever is, is." And since Locke believes, with Descartes, that we are conscious of all that passes through our minds, if children did believe such maxims they would be aware of these beliefs. Since they aren't, we have evidence that these principles are not innate. More formally and using the principle of noncontradiction – that it is impossible for the same thing to be and not be – as an example of an innate principle the argument is:

> (1) If the principle of noncontradiction is innate then young children will believe that it is impossible for the same thing to be and not be.
> (2) If young children believe that it is impossible for the same thing to be and not be, then they will be aware that they believe it.
> (3) Young children are not aware that they have the belief that it is impossible for the same thing to be and not be.
> (4) Thus, young children don't believe the principle of noncontradiction.
> (5) Thus, the principle of noncontradiction is not innate.

Premise 1 follows from Locke's understanding of innateness, which is discussed in more detail below. Locke believes that premise 3 is empirically supported, and I'm going to assume that he is correct, because I want to focus on premise 2.[7] The second premise is a special instance of his general commitment to the claim that if we have an idea or know a truth we are necessarily conscious of that idea or truth. This is sometimes called the *transparency thesis*. If the transparency thesis is false, then the

[7] How to interpret the empirical evidence is a contentious issue. There are reasons for denying 3, but I'm not going to address these worries. It seems to me that the philosophically crucial issue is how to understand innateness; until the conceptual difficulties are cleared it is difficult to assess the evidence for 3.

fact that children don't appear to have such principles or ideas is no evidence against their possessing them.[8]

Locke's support for the transparency thesis comes from his understanding of what counts as having an idea. According to Locke, the notion that an idea could be imprinted on a subject's mind while the subject remains unaware of the idea is "hardly intelligible" (Bk I.Ch.1.5). In the charge of unintelligibility we see beginnings of the second horn of the dilemma: the claim that innate ideas do no explanatory work. After stating that the nativist claim is unintelligible, Locke goes on to say, "To say a notion is imprinted on the mind, and yet at the same time to say, that the mind is ignorant of it, and never yet took notice of it, is to make the impression nothing" (Bk I.Ch 1.5). In other words, in order for the theory of innate ideas to be considered a *theory*, the innate ideas must provide an explanation of some data. If innate ideas are merely items in the mind that we cannot access until we access them, i.e. until we become conscious of them, then the innateness of the ideas is explanatorily inert.

The force of the second horn of the dilemma becomes clearer when Locke considers the rejoinder to his claim that children don't believe innate principles: that while young children don't assent to these maxims, once children reach the age of reason they readily give their assent. He first argues that this is empirically false; children begin reasoning before they begin to assent to such principles, and many adults never give their assent (Bk I.Ch. 1.12). But, he claims, even granting that everyone comes to assent to the maxims once they begin to reason would not prove that they are innate, unless we are willing to claim that *all* propositions which require the use of reason to gain assent are innate. These would include not just the self-evident maxims like "whatever is, is", but all the truths that are incomprehensible until

[8] Since Descartes himself championed the transparency thesis, Locke begs no question against him by employing it in the above argument.

we apply our rational faculties. That is, there would be no way to distinguish innate ideas from any other ideas which might arise after the rational faculties mature. And if there is no way to distinguish these innate ideas or principles from those that arise through the operations of our mental faculties, then they are superfluous to an explanation of how our minds work.

Of course, the significance of Locke's dilemma depends on how we understand the account of his adversaries, the nativists. A common complaint against Locke is that he has attacked a straw man.[9] It's claimed that the version of nativism that he challenges is so crude that no philosopher worth his stripes would hold it. In the remainder of this section, I distinguish three versions of nativism and argue that Locke is not guilty of attacking a straw man.

There are many different ways to work out a theory of innate ideas, but any theory of innate ideas needs to explain (i) the nature of these ideas and (ii) the process by which we become aware of these ideas. In the following discussion, I want to draw your attention to how the first account differs from the second and third with respect to (i), and how the second and third accounts differ in their response to (ii).

First to be discussed is the *crude account* of nativism. On this account, (i) we are born with fully formed ideas of God, substance, etc. and (ii) we are aware of these ideas from the beginning. Innate ideas are understood as either images or definitions, and they exist in the mind complete at birth. This crude account is illustrated by the simile between a craftsman and God: just as a potter marks his work with a symbol to indicate its authorship, so God imprints the mind with (for example) an idea of Him. Neither part of this account is especially plausible. It seems that our ideas of God or substance are not fully formed; the details of how we think about God or substance

[9] The Fraser edition of the Essay, which I am using, is full of footnotes complaining about Locke's crude understanding of nativism.

are informed by our particular culture and upbringing. As for (ii), there is little reason to think that newborns have a second-order awareness of their own ideas.

A more sophisticated version of nativism is the *propensity account*. Innate ideas aren't images or definitions imprinted on the mind, but consist in tendencies to respond in certain ways. Some of the most famous proponents of nativism adopt something like the propensity account. Leibniz, for example, compares innate ideas to veins in a block of marble which outline a particular shape, e.g. Hercules, in contrast to a block of marble without veins marking any particular shape: "the block would be more determined with respect to that shape and Hercules would be as though innate in it in some sense" (*Preface to the New Essays*, p. 53). Leibniz goes on to say that just as the sculptor must labor to clear away the excess marble, people must work in order for the innate ideas to appear. The block contains a tendency towards a Herculean shape, but the details of the finished statue will depend on the sculptor (and the larger environment). Similarly, all people are born with a tendency to believe in God, but precisely how that belief is manifested will depend on the environment.[10]

Descartes also understands innateness as a propensity to acquire certain kinds of ideas. In an effort to distinguish innate ideas from other ideas, he compares them to the predisposition to manifest a character trait or acquire a disease:

> This is the same sense as that in which we say that generosity is 'innate' in certain families, or that certain diseases such as gout or stones are innate in others: it is not so much that the babies of such families suffer from these in their mother's womb, but simply that they are born with a certain 'faculty' or tendency to contract them. (*Comments on a Certain Broadsheet*, 215).[11]

[10] This basic idea has gained recent favor among certain scholars interested in the evolutionary psychology of the near universal human belief in higher powers of one sort or another.
[11] Page numbers refer to *The Philosophical Writings of Descartes* (1984) trans. Cottingham.

The propensity account avoids certain objections. Unlike the crude account mentioned above, Descartes denies that the ideas exist in a child's mind while still in the womb. Just as the environmental conditions need to be right for gout to strike, so too must the conditions be right for us to be aware of our tendency to assent to a particular maxim or acquire a certain idea. The propensity account explains why children don't assent to innate principles; it's plausible to think that children must reach a level of rational maturity before their innate ideas are manifest. The propensity view avoids another objection as well: not everyone in a gout-stricken family will get gout, just as not everyone with the innate tendencies will manifest them. Sometimes the conditions won't ever be right, and the tendency will never be manifest.

Finally, there is the *knowledge account* of innateness. The knowledge account begins with the claim that people have a tendency to assent to certain universal propositions or acquire an idea, and then explains the possession of these tendencies by appealing to innate knowledge. The universal assent to the principle of noncontradiction is explained by the fact that we are born with knowledge of this very proposition. The universal acquisition of the idea of substance is explained by the fact that we are born with knowledge of substances. This version of nativism holds that evidence for innate ideas are the tendencies we have, but what explains these tendencies is some prior knowledge, knowledge which is present from birth, but that we are not aware of until certain conditions have been met.

The propensity and knowledge accounts both describe innateness as a tendency to acquire an idea as long as certain conditions are met. So the two accounts agree on criterion (i) the nature of innate ideas. But the two accounts differ in their response to criterion (ii) the explanation of the process by which subjects become aware of these ideas. On the propensity account subjects become aware of these tendencies, i.e. the

tendencies become manifest, when the subject reaches a certain level of cognitive development. The propensity account offers an explanation which is basically a description of the psychological facts. Thus far, the knowledge account is in agreement with the propensity account. But the knowledge account goes one step further. According to it, the *existence* of these tendencies is explained in terms of a prior innate knowledge.

Only the knowledge account of innateness would be a rival to empiricism as a theory of knowledge, because only this version offers an explanation for how we know certain propositions. [12] Locke can agree with the nativist, and does, that we are born with certain mental capacities. So, in this limited sense, he can agree that we have innate tendencies to assent to propositions or acquire certain ideas. The fact that we are born with certain mental faculties (or, at the least, the tendency to develop certain faculties) is not in dispute, nor is the fact that these faculties are necessary for understanding certain truths. Rather, Locke's complaint is that the nativist gives us no way to distinguish an innate truth from a truth acquired on the basis of experience (Bk I.Ch. 1.5). Both the nativist and the empiricist have a story about the source of the tendencies: the nativist claims that the source is innate knowledge and the empiricist claims that the source is experiential learning. But if, as Locke argues, the two categories cannot be distinguished, then it seems to be a distinction without a difference: thus, either ideas are all innate or all experiential.

It's worth pointing out that Descartes, at one time, appears to have endorsed something close to the claim that all our ideas are innate. Descartes thinks that our sense organs are stimulated by different kinds of movements in the nerves (*Optics*, p. 64). Both the particular movement of the nerves and the particular pathway the movement takes place in determine what kind of sensation is present in the soul. So,

[12] Cf. Atherton (1983, pp. 54-55).

for example, a particular type of movement in the nerves of the tongue will cause a bitter taste sensation. Descartes also thinks, quite reasonably, that there is no resemblance between the movements of the nerves and the ideas in the soul. This leads him to claim, "The ideas of pain, colours, sounds, and the like must be all the more innate if, on the occasion of certain corporeal motions, our mind is to be capable of representing them to itself, for there is no similarity between these ideas and corporeal motions" (*Comments on a Certain Broadsheet*, 216). If what Descartes intends by this claim is that our minds come equipped with the ability to represent physical stimuli from the external world by innate categories of sensations, then Locke would not disagree. He, too, accepted that the brain is stimulated by corporeal movement, and that we all have the same basic sensations.[13] If the claim that we are born with the capacity to see colors, hear sounds, etc., is all that Descartes means by "innate ideas" then the claim that we are born with innate ideas is not in dispute.

But I think there is a dispute, and the dispute is in part methodological. Descartes begins by assuming that there are truths which we know with certainty, e.g. pain hurts, and goes on to work out what conditions must be like in order for us to possess this knowledge. Since the experience of feeling pain does not in any way resemble the corporeal movements which correlate with it, it must be that the idea of pain is innate. Locke, on the other hand, starts by looking at the mental faculties people have, and investigates these abilities to figure out what we can know; his theorizing is driven by his observations.[14]

---

[13] Locke's comparison of the mind to a blank slate is misleading in its suggestion of total passivity; in his desire to distinguish his view from the nativist Locke overstates the case. The Lockean conception of mind includes congenital faculties that allow us to learn from our experiences. Some of these basic faculties include: perception, retention, discernment (the ability to distinguish one thing from another), comparison, and abstraction. These faculties perform operations on the ideas provided by experience, and contribute to the nature of our perceptual experience.

[14] Don Garrett (1997, pp. 30-33) calls this "methodological empiricism".

To return to the dilemma that faces the proponent of innate ideas: either innate ideas are understood in such a way that it turns out that it is empirically false that they exist, or they are understood in such a way that their existence turns out to be trivially true. Understanding innate ideas in the crude way, as full-blown images or definitions which we are aware of from birth, runs up against the first horn of the dilemma. No such ideas exist.[15] Understanding innate ideas as propensities, version two above, bypasses the first horn. Since propensities require that certain conditions be present to be triggered, the fact that children and some adults fail to manifest the tendency does not show that such tendencies do not exist; an alternative explanation is that some conditions have failed to hold. But the propensity account bumps up against the second horn. If we understand that the tendencies consist merely in the assent to certain propositions once rational maturity has been reached, then there is no reason to distinguish the innate truths from any other truths reached with the aid of reason. The truths' innateness no longer plays any role in explaining their presence in a rational mind, since the innate truths would be assented to whether or not they were innate. Thus, either all the truths based in reason are innate, or the innateness of the truths is superfluous to the explanation of how we know them. Locke doesn't take the view that all the truths acquired through reason are innate as a serious option. Thus, we are left with the conclusion that the innateness of certain principle and ideas are superfluous when it comes to giving an explanation of the knowledge.

On this interpretation, the transparency thesis is crucial for both horns of Locke's dilemma.[16] As for the first horn, premise 2 of the argument for the conclusion that the principle of noncontradiction is not innate *is* the transparency thesis. And the

[15] As mentioned in fn. 7, the issue is not so simple. The debate between nativists and empiricists is ongoing. See Spelke (1998) for a discussion of the empirical evidence that favors the nativist. See Prinz (2002) for a defense of empiricism and Fodor (1998) for a defense of a radical nativism.
[16] Atherton (1983, p. 51).

transparency thesis underlies Locke's frequent complaints that we could be imprinted with a truth but be unaware of it is "hardly intelligible," which constitute the second horn of the dilemma: "For that a truth should be innate and yet not assented to, is to me as unintelligible as for a man to know a truth and be ignorant of it at the same time" (Bk I. Ch. 1. 24; see also Bk I.Ch 1.5).

The claim of unintelligibility can sound disingenuous. Even if, as argued above, there seems to be no reason to *accept* the claim that we have innate knowledge we aren't aware of, the claim itself seems comprehensible. We can imagine, for example, a world in which people are born with knowledge of certain truths and the learning process consists in trying to recall or uncover these truths. (The claim that we all have past lives, and we can uncover memories of our earlier lives by a process of recollection, also seems conceptually possible.) There doesn't seem to be anything contradictory or incoherent about such a theory of knowledge unless you are assuming the transparency thesis. Locke's belief that such a situation is unintelligible follows from the transparency thesis; for Locke, it's part of the concept of "idea" that we are always aware of having them.

Why does Locke commit himself to the transparency thesis? It's possible that it just never occurred to him to question it, but I think there is another possible explanation. I think Locke uses the transparency to provide the limits of his inquiry; in other words, the transparency thesis clearly distinguishes the mental from the non-mental. Unlike Descartes (and later Leibniz), Locke doesn't attempt to explain the connection between the mind and the body, and explicitly limits his area of inquiry to the mental:

> I shall not at present meddle with the physical considerations of the mind; or trouble myself to examine wherein the essence consists; or by what spirits or alterations of our bodies we come to have any *sensation* by our organs, or any ideas in our understandings; and whether those

> ideas do in their formation, any or all of them, depend on matter or not. (Bk I.Ch. 1. 2)

For anyone who does not identify mental states with brain states, there needs to be a way to distinguish the mind from the brain. Typically, philosophers go one of two ways: either the mental is marked by its intentional character or by its phenomenal or qualitative character. Locke accepts the latter marker, and adds that a state cannot be phenomenally conscious without the subject's awareness. While the view that a state's being conscious requires that we are aware of being in that state strikes many contemporary philosophers as false, Locke doesn't come to this belief without consideration.[17] He discusses, for example, a situation where a person is distracted and thus fails to notice some noise in the background.[18] Locke claims that in such a situation, despite the fact that there is the usual motion in the ear, there is no perception. Only when we become aware of the music do we have a sensation: "wherever there is sense perception, there is some idea actually produced, and present in the understanding" (II.9.4).[19] For Locke our awareness of our ideas distinguishes them from the physical effects in the body that, in some way he doesn't try to understand and perhaps thinks is beyond the ability to understand, correspond to the ideas.

Locke needs a way to distinguish the mental from the material which will in turn mark the limits of his inquiry. To say that an idea can exist in the mind unperceived is to blur the line between the physical and the mental, since it would allow that the physical states of the brain, like the movement of the sound waves in the ear, count as unperceived ideas. Moreover, while commitment to the transparency

[17] It doesn't strike all contemporary philosophers as false. In Chapter Four, I discuss William Lycan's (1996) higher-order account of consciousness, which traces its roots back to Locke's account of the role of inner sense.
[18] This type of case is familiar to contemporary philosophers from the distinction between access and phenomenal consciousness. See Block (1997).
[19] What would Locke say about blindsight?

thesis provides a boundary for an inquiry into the mental, it also promotes introspection as a primary source of information about the mind. If Locke is skeptical about our abilities to understand the mind-body connection, he is perhaps too optimistic about our ability to understand our own selves. It follows from the transparency thesis that all that is needed for us to understand our own minds is careful introspection-- it is all there for us to directly perceive.

2.1.2 Locke's Two Thought Experiments

Locke proposes his empiricist theory as an alternative to the nativist explanation of how we acquire ideas. If our ideas aren't present in our minds to begin with, then they must come from somewhere. Locke believes that all ideas, which he defines as the objects of thought, arise from one of two sources, sensation or reflection. Ideas of sensations, like whiteness, bitterness, softness, or warmth, are brought to the mind through the sense organs (Bk II. Ch. 1. 3). Ideas of reflection, which include perception, doubt, and thought, arise from attending to the operations of our minds (Bk II. Ch. 1. 4). All of our ideas, i.e. all of the objects of our thoughts, are the product of one of these two sources. (As mentioned above, the mental faculties themselves Locke thinks we are born with.)

Locke distinguishes among many different kinds of ideas, but most central to his empiricist theory of knowledge acquisition is the distinction between simple and complex ideas. Though in the world sensible qualities like colors, firmness, movement, warmth are all "so united that there is no separation, no distance between them" in the mind the idea of each quality is simple and unmixed (Bk II, Ch 2, 1). What distinguishes simple from complex ideas is that the former have a "uniform appearance;" simple ideas are those which cannot be further analyzed on the basis of

appearance, and which can be traced back to a particular sensation or reflection.[20] Moreover, unlike complex ideas, simple ideas can be neither created nor destroyed by the mind.

Complex ideas can be both created and destroyed by changing the arrangements of the simple ideas. You can imagine, for example, a creature with the shape of a pig's head and torso, the color of a daffodil, the wings of a turkey and the bray of a donkey. Thus, we create a new complex idea. But, Locke claims, no one, no matter how smart or creative, can "invent or frame" a new simple idea, that is, a simple idea that has not come from the processes of sensation or reflection (Bk II. Ch. 2. 2). Locke's argument is again empirical. He challenges the reader to think of a new simple idea, and points out that successfully meeting this challenge would be equivalent to a blind man having ideas of colors, or a deaf man "true distinct notions" of sound.[21]

We're now in a position to investigate the positive side of Locke's claim that all the materials of the mind are furnished by experience. We've seen that on one reading, the empiricist principle is merely a rejection of the view that there are innate ideas. But the empiricist principle is not only a rejection of innate ideas; it is also a theoretical claim in its own right, which can be understood in different ways. As I noted at the beginning of the chapter, the empiricist principle is primarily a claim about the source of our ideas and beliefs.

Locke's understanding of the empiricist principle is weak compared to Hume's, which is discussed below. Locke, *pace* Hume, does not think that there is a

---

[20] Of course, if you are perceiving a uniformly colored swath you could analyze it into smaller pieces. The point is that you cannot further break down the color appearance.

[21] By qualifying the word "notion" Locke implies that the deaf man can have a distorted or unclear idea of sound despite never having heard any. Perhaps this is inadvertent, or perhaps Locke's intuitions here are not as strong as the color case. It is easier to put yourself in the situation of a blind person—you cover your eyes or sit in a dark room—than it is to put yourself in the situation of a deaf person. It's much more difficult to eliminate sounds.

one to one correspondence between our ideas and the experience (impressions) that suggest them. Locke thinks that some of our ideas come from a variety of sources. Locke distinguishes between simple ideas that can be conveyed by only one sense and those which can be conveyed by more than one sense. In the former category are simple ideas of colors, noises, tastes, smells, and heat (Bk II. Ch. 3. 1). In the latter category are the simple ideas we have of extension, shape, and movement (Bk. II, Ch. 5). The blind have no idea of colors, but they do have ideas of shapes since shapes can be both seen and felt.

To defend the claim that all our ideas arise from experience of either the external world or the internal one, Locke again refers to children's development. Most of our ideas are conveyed to our mind when we are children, but in cases where we fail to have the requisite experiences we fail to acquire the ideas:

> Light and colors are busy at hand everywhere when the eye is but open; sounds and some tangible qualities do not fail to solicit their proper senses and force an entrance to the mind; but yet, I think, it will be granted easily that if a child were kept in a place where he never saw but black and white until he were a man, he would have no more *ideas* of scarlet or green than he who from his childhood never tasted an oyster or a pineapple has of those particular relishes. (Bk. II, Ch. 1. 6)

To support his thesis that possessing an idea requires having the requisite experience Locke begins with an extraordinary case, which he uses to elicit intuitions that support more mundane instances of the thesis. Foreshadowing the knowledge argument, he imagines a boy locked away in a black and white room who never sees any colors. Clearly, Locke says, this boy will not have any ideas of red or green. For precisely the same reason, a man who has never tasted pineapple will lack the idea of the taste of pineapple. (It's worth noting that on the face of it there seems to be difference between a boy who has never seen colors, and man who, though having many gustatory experiences, has never tasted pineapple. On the first case, there is an entire

category of experience types missing, while in the latter, there is only one particular type of experience missing. As we'll see, Hume also argues from the general category case to the particular experience type without comment.)

And like another famous twentieth century argument, Thomas Nagel's argument that we cannot know what it is like to be a bat, Locke argues that although it is merely contingent that we have five primary senses—we could have had fewer or we could have had more—it is impossible for us to imagine other kinds of sensible qualities (Bk II. Ch. 2. 3).[22] Locke suggests that it's likely there are creatures somewhere in the universe, who have different types of sense organs than our own, and thus have ideas of sensible qualities which do not correspond to our own. This argument seems to be an extension of the case of the boy raised in a black and white environment. Just as he is unable to form an idea of colors, we are unable to form an idea of sensible qualities available to alien creatures.

This argument is more complicated than the boy in the black and white room, because it can *appear* to be a counterexample to Locke's claim that we can only have ideas of things we have experienced. We are asked to suppose that we could have had more sense organs than we currently have; that is, we could have had different ideas of the sensible qualities of objects. But, according to the empiricist principle, we cannot imagine what these qualities could be, since we've not had the right sort of experience. Here is the problem: to understand Locke's argument we must be able to have an idea of sensible qualities which we have never experienced. We must be able to understand the proposition that it's possible that there are sensible qualities, which are impossible for us to imagine. But this proposition includes the idea of sensible qualities which we have never experienced (in fact, cannot ever experience), which is impossible according to the principle that we must have the experience to get the idea.

[22] See Nagel (1974).

If the two cases are of the same type then just as the boy cannot understand, because he lacks the ideas, a proposition that involves the idea scarlet, we should be unable to understand a proposition that includes the idea of a sensible quality which we have never experienced. There seems to be tension between the implications of the empiricist principle and the ability to imagine the alien case. This would be bad news for the empiricist principle, since it seems obvious that we can understand the alien case.

The apparent difference between these two thought experiments is partially caused by the fact that they are made from different points of view. In the case of the boy, we are imagining the situation from the third person, and in the case of the alien we are imagining the situation from the first person. A second real difference is what we are asking. In the first case, we ask ourselves whether the boy has an idea of color and answer negatively; we are not concerned with what ideas the boy *does* have. The boy is going to lack the general idea of colors, since he has never seen *any* particular colors from which he could form a general idea. But he does, of course, have the general idea of sensible qualities, since all he lacks are color experiences. He is in the same position towards us that we are towards the alien creatures. He can know that we have sensible ideas which he cannot imagine.

If the boy can understand propositions that refer to experiences he has never had, what is it that he fails to know? It's interesting to compare Locke with Nagel here. Nagel famously argued that what we fail to know about the bat is *what it is like to be a bat*. Nagel is assuming that the bat has conscious experiential states and that these states have a distinguishing qualitative feel. When the bat is flying through the night sky positioning itself in space through echolocation there is a subjective feeling that characterizes what it is like to be that bat at the particular time. Moreover, this subjective feeling is one that is familiar to any normal bat. (It is a type rather than

token property.) According to Nagel, we cannot imagine what it is like to be a bat because we lack the conceptual framework that determines what experiences are like for bats.[23]

A contemporary reader of Locke might naturally assume that something similar is being proposed in his thought experiment about the alien creatures. It's impossible for us to know the alien sensible qualities because it is impossible for us to imagine what experiences the alien creatures are undergoing. But, in fact, that does not seem to be what Locke had in mind. Locke argues that we can't imagine is not the alien ideas, but the alien sensible qualities of the objects: "it is not possible for any *man* to imagine any other qualities in bodies, howsoever constituted, whereby they can be taken notice of" (Bk II. Ch. 2. 3). What the boy can't imagine while stuck in his black and white room is a property of objects, namely what colored objects look like. What we can't imagine about alien creatures is how objects appear to them. What we fail to know, according to Locke, is a claim about a property of objects rather than a property of our experience of objects.[24]

Conclusion

In this section, I articulated and discussed the empiricist principle endorsed by Locke. I've shown that Locke has two kinds of arguments for the empiricist principle, negative and positive. Locke's arguments against the nativist do not follow from his commitment to empiricism, but stand on their own. He presents the nativist with a dilemma: either it is empirically false that innate ideas exist or they are indistinguishable from ideas acceptable to the empiricist. I argued that Locke's

[23] At least, this is one of Nagel's arguments. There seems to be a different argument for the same conclusion, which relies on the claim that we can never inhabit more than one perspective at a time. See remarks on pp. 339-343.

[24] In this, Locke is in line with some contemporary philosophers in thinking that we should understand that what Mary learns when she sees red for the first time is what red objects look like, and not what it is like to have the experience of red. This is discussed in Chapter Four.

arguments do not rely on a straw man version of nativism, but do require that nativism be understood as presenting an explanation of the allegedly innate knowledge. When understood as providing a theory of knowledge, according to Locke, nativism fails. Locke's arguments against the nativist rely on the transparency thesis, the claim that we are aware of all that passes through our minds. I offered a suggestion for why Locke might have found such a claim compelling, but recognize that this will not be persuasive to most contemporary philosophers.

I then investigated Locke's arguments in support of the empiricist principle. I concluded that the difficulty in fleshing out the empiricist principle is that Locke's understanding of the nature of perceptual experience is underdeveloped. I described two thought experiments that Locke offers in favor of the empiricist principle, the boy in the black and white room and the aliens with different sense organs. While these are of historical interest as forerunners to influential thought experiments of the latter half of the twentieth century, Locke's arguments do not go beyond pumping intuitions.

2. 2 The Nature of Perceptual Experience

I now turn to the second thought experiment I wish to discuss, Molyneux's Question. Molyneux's Question is of interest for many reasons. First, Locke, Leibniz, and Berkeley all discuss the thought experiment and it is interesting to watch the historical development of the empiricist responses. Leibniz responds directly to Locke and Berkeley follows up on Leibniz's criticism by offering a more mature theory of perceptual experience. Second, the discussion in this section moves us away from Locke's version of the empiricist principle, which emphasizes the *origin* of simple ideas, to Berkeley's concern with providing a theory of the nature of these ideas. (In the final section of this chapter, we'll discuss Hume's version of the empiricist principle which combines parts of both Locke's and Berkeley's views.)

2.2.1 Molyneux's Question and Locke's Reply

William Molyneux, whose wife was blind, posed the following question to Locke in a letter dated March 2, 1693, which Locke inserted into the second edition of the *Essay*. Molyneux asks:

> Suppose a man born blind, and now adult, and taught by his touch to distinguish between a cube and a sphere of the same metal, and nighly of the same bigness, so as to tell, when he felt one and the other, which is the cube, which is the sphere. Suppose then the cube and sphere placed on a table, and the blind man be made to see: *quaere*, whether by his sight, before he touched them, he could now distinguish and tell which is the globe, which the cube? (Bk II. Ch. 9. 8)

Molyneux's question has been the subject of much debate; blindness was a popular topic during the modern period, and medical advances in the 18th century, which allowed for the removal of cataracts, thus bringing sight to the blind, served to fuel this interest.[25] And Molyneux's Question continues to be of interest to contemporary researchers.[26] Molyneux answered his own question with an unequivocal "not," and Locke agrees. But before examining Locke's response, I want to distinguish among three possible interpretations of the question, each of which assumes that the previous question can be answered positively.

The first interpretation of Molyneux's Question is popular with psychologists and visual scientists.[27] They frequently take the question to be the following: would the blind man have enough of a functioning visual apparatus, both the physiological

[25] For the history of Molyneux's Question see Chapter 2 in Riskin (2002) and Morgan (1977). One of the most famous accounts of the "couching" of cataracts in a young boy is William Cheselden's (1727-8). The relevance of the empirical studies to the philosophical questions raised by Molyneux's Question is discussed in more detail below.

[26] See Gallagher (2005) for a contemporary, empirically informed response to Molyneux's Question. Gallagher argues, on neurological grounds, that the newly sighted blind man wouldn't be able to identify the cube and the sphere.

[27] It's worth noting that this question was raised early on, by Denis Diderot (1749). Translations of large portions of Diderot and Condillac are found in Morgan (1977).

and the cognitive elements, to be able to see the cube and the sphere at all upon first opening his eyes?

There is a second, related, question: would the blind man be able to distinguish the cube from the square? In other words, would the cube and sphere appear to the newly sighted to be different kinds of objects?

Finally, there is the third question of whether the blind man could correctly *identify* the cube and the sphere. The blind man understands, from his haptic shape concepts, the words "sphere" and "cube." But would the newly sighted blind man correctly categorize the two shapes based on vision alone? Could he subsume the shapes under his haptic concepts?

As I just mentioned, each of these questions assumes the answer to the preceding questions is "yes." If the blind man cannot see the objects, then he won't be able to distinguish them. If he can't distinguish the two objects, then he won't be in a position to identify one as the cube and one as the square. In his response, Locke focuses on the third question: "the blind man, at first sight, would not be able with certainty to say which was the globe, which the cube, whilst he only saw them" (Bk II. Ch. 9. 8). Locke answers the third question in the negative, and assumes that the blind man would be able to see the two shapes.

Later writers have criticized Locke for making this assumption.[28] But I don't think this criticism is fair. (Moreover, I think it can be misleading as I discuss directly below.) The most obvious defense of Locke's decision to ignore the first two questions is that the most philosophically interesting question is the third. For the purpose of isolating the third question it's reasonable to assume that the blind man could both see and distinguish the cube and the sphere. More importantly, we

[28] Diderot and Condillac both criticize Locke for assuming that the blind man would be able to see immediately upon opening his eyes.

shouldn't take the fact that Locke grants the newly sighted blind man the abilities to both see and visually distinguish the cube and the square to be evidence that he actually believes that the blind man would have these abilities immediately upon gaining sight.

The idea that our sense experiences come to us fully formed has led to charges that for Locke the mind is completely passive. Combined with his metaphors of blank slates and empty cabinets, these kinds of remarks inadvertently distort and simplify Locke's actual account of how the mind works. How we understand the passivity of the mind will affect how we understand Locke's conception of experience. The blank slate metaphor and the talk of "imprinting" makes it sound like the ideas come into the mind fully formed, and mark the mind like a rubber stamp on a piece of paper. But this is clearly not how we should understand the Lockean mind. There are passages which make clear that Locke did not think that the mind waits quietly, allowing itself to be imprinted by sensations from the outside world. Even in perception, the most passive of processes, the mind contributes to how the world appears to us (see Bk II. Ch. 9 on the faculty of perception).

Of course, in one sense, the mind is completely passive when it perceives: we open our eyes and we cannot help but see what is in front of us. We cannot will ourselves to see, no matter how fervent our desires, what is not there.[29] But the disagreement between the Rationalists and the Empiricists was not over this sort of passivity—both can agree that the mind is passive in the sense *that* we perceive automatically when our senses are working properly.

But *what* we perceive is affected by the operations of our minds. Locke believes that the visual images we receive are two dimensional; when we look at a

[29] Of course, we do sometimes see things that aren't there. My point is that we cannot will ourselves into having a full-blown perceptual experience, veridical or nonveridical. If you are looking at a red tulip, you can't will yourself to perceive it as yellow.

round globe "the idea thereby imprinted on our mind is of a flat circle, variously shadowed, with several degrees of brightness coming to our eyes" (Bk II. Ch. 9. 8).[30] But, as adults, we perceive a world of objects that appear three dimensional and uniformly colored. Locke thinks that we have learned, through repeated experiences, to change the ways objects appear. As he says, "the ideas we receive by sensation are often, in grown people, altered by the judgment, without our taking notice of it." The ideas are changed by the repeated experiences of material objects, until, as the result of habit, "the judgment…alters the appearance into their causes" (Bk II.Ch. 9. 8). Sensible ideas do not arrive in the mind fully formed, but are shaped by past experiences; we have to learn how to see.

Given Locke's views about the effects of repeated experience, it seems likely that he would have denied that the blind man would have been able to do much of anything upon first opening his eyes. Locke claims that his reason for relating Molyneux's question to the reader of the *Essay* that he wants the reader to recognize "how much he may be beholden to experience, improvement, and acquired notions." This passage, in particular its mention of the need for improvement, suggests that Locke very well might have denied the two assumptions necessary for posing the third question.

But Locke is primarily concerned with the question of whether the blind man could correctly identify the cube and the sphere. Molyneux claims that the blind man will lack the relevant knowledge because he will fail to immediately associate the visual concepts he receives upon regaining his sight with the haptic concepts he already has: the blind man "has not yet obtained the experience, that what affects his

[30] Later in the passage, Locke points out that creating three dimensional images out of the two-dimensional nature of our visual experience is what we do when looking at a painting. Berkeley also believed that the visual system provides us with two-dimensional images. See Schwitzgebel (2006) for an interesting, historically informed discussion of the claim that vision is two-dimensional.

touch so or so, must affect his sight so or so" (Bk II. Ch. 9. 8). Locke agrees with Molyneux that the blind man would be missing this ability. But it doesn't follow from Locke's empiricism that the blind man would be unable to reason from his haptic idea to his visual one. Locke, unlike Berkeley and Hume, does not think that there is a one to one correspondence between sensations and ideas. As discussed above, there are ideas that Locke thinks are the result of different senses, for example, the ideas of existence and space (Bk II. Ch. 5. 1). It seems that this would also be possible for ideas of shapes; the same shape idea might be formed by both the haptic and visual ideas in sighted people.

In fact, this process of concept formation is especially congenial to Locke's account. As we've just seen, Locke thinks that the visual images we have of the world are two-dimensional; through vision, the round globe is imprinted on the mind as a flat circle. But repeated experience changes the way we see objects: "habitual custom, alters the appearances into their causes" (Bk. II. Ch. 9. 8). One way we can learn that what appears as a flat circle is actually a round globe is to walk around it, and see the object from a variety of different perspectives. But another, more efficient, way to learn that the flat circle is a round globe is to pick it up with our hands. If customs which alter the visual appearances of objects into their causes include customs arising from the sense of touch, then it seems fair to attribute the knowledge of three-dimensional shapes to both senses.

And if this account of concept formation were correct, then it would seem reasonable to suppose that a newly sighted man, who we assume can see accurately and clearly from the beginning, will be able to reason from the haptic idea he already possessed to the visual idea he currently has. It seems plausible that he could take visual note of the pointy corners of the cube and reflect on his tactile feelings of sharpness when handling the cube, which would allow him to infer that the object on

the left must be a cube. The subject of Molyneux's question is an adult, with fully developed cognitive powers. If in the ordinary case a person forms her three-dimensional shape concepts through more than one sense, then why couldn't the Molyneux subject reason from one sense to another?

Locke does not consider this prediction, even though it seems to be consistent with his commitments. Why not? Perhaps he does not consider this prediction because he takes his opponent to be a nativist. The discussion of Molyneux's question follows a discussion of innateness in which Locke admits that children are born with certain ideas, namely those of hunger and warmth. But these ideas are the result of ideas that are formed in response to conditions of the womb. So these are not problematic innate ideas, but ideas formed under the same kind of conditions that hold during the formation of all the other sensible ideas. Moreover, given the large amount of space devoted to the issues of innateness in the *Essays*, I think it is natural to assume that in responding to Molyneux Locke had in mind the nativist.

What would the nativist position on Molyneux's question be? Assuming that the nativist includes shapes among the innate ideas, it seems that the newly sighted should be able to correctly name the shapes; that is, the nativist should respond positively to Molyneux's question. Assuming, as I argued above, that the only innate theory that counts as a genuine rival to the empiricist principle is the one that claims that our innate knowledge explains our dispositions, we can see why Locke moved so quickly through his answer to Molyneux. He can apply the same dilemma here as he used against innate ideas generally. Either the nativist is empirically wrong or the claim that the blind man would correctly name the cube and square is trivially true. In Locke's time, the evidence wasn't available, so on the empirical question Locke and the nativist are at a stalemate. But to avoid the second horn the nativist would be forced to conclude that *immediately* upon gaining sight the formerly blind man should

be able to see. If, instead, the nativist allows that it might take some time for the blind man to correctly identify the objects, then her theory of innate ideas has collapsed into a mere dispositional claim, which is only trivially true.

2.2.2 The Geometrical Account and Berkeley

Those who answer Molyneux's Question with a "yes" do not always explain their position in terms of innate ideas. Instead, they frequently make a distinction between having an *image* of a thing and having an *idea* or conception of a thing. Leibniz, for example, thought having an idea of a thing is like having a definition.[31] In discussing Molyneux's Question, Leibniz claims that the blind man will be able to reason along the lines mentioned above. His tactile image of a sphere differs from his tactile image of a cube in that the former has no points whereas the latter has eight. From the tactile image, he includes in his *idea* of a sphere the condition that it has no sharp points. So, when he is confronted with visual images of a sphere and a cube, the blind man's idea of a sphere will allow him to correctly name the sphere on the basis of vision.

Leibniz' response does not depend on the existence of innate ideas. Instead, it relies on a distinction between the images given to us by the different senses and the ideas which we form on the basis of such images. Leibniz conceives of these ideas as being definitions, definitions that are abstracted away from the images which are their source. He argues that there must be ideas like this, since the blind and the paralyzed are both capable of understanding geometry in the same way, despite the differences in how they acquired the geometrical ideas: "These two geometries, the blind man's and the paralytic's, must come together, and agree, and indeed ultimately rest on the same ideas, even though they have no images in common" (2.2.8).

[31] Leibniz' discussion of Molyneux's Question is in II.9.8 of *New Essays on Human Understanding*, his commentary on Locke.

By basing his response on the distinction between ideas and images, Leibniz changes the nature of Molyneux's question. I've understood Locke as interpreting the question as one about the source of our ideas, but Leibniz, despite his interest in defending innate ideas, investigates the nature of ideas generally. For Leibniz the issue that arises from Molyneux's thought experiment is how to understand the nature of ideas, in particular abstract ideas. As Leibniz recognizes, his answer to Molyneux's question seems compatible with Locke's views; that is, it is compatible with Locke's views to answer "yes" to Molyneux's question.[32]

Berkeley, however, offers a thoroughly empiricist response to Molyneux in *An Essay Towards a New Theory of Vision* (1709). Like Leibniz, Berkeley assumes that the significant issue raised by Molyneux is the nature of ideas. But he disagrees with Leibniz' views on the natural geometry of vision, that is, with Leibniz' rationalist conception of experience. Berkeley's response to Molyneux follows from his views on the nature of perception, in particular the nature of vision and touch and the relationship between these two senses.

What made Berkeley's theory of vision "new"? Primarily, it is the theory's empiricist explanation of the ability to see distances. At the time of Berkeley's writing, it was commonly believed that in vision we do not immediately perceive distances or magnitudes. Berkeley, rather uncritically, accepts this view: "It is, I think, agreed by all, that *distance* of itself, and immediately, cannot be seen" (II).[33] Where Berkeley differs from the received view is in his explanation of how we come

[32] See his comment in section 136: "It may be that Mr. Molyneux and the author of the Essay are not so far from my opinion as at first appears."

[33] Berkeley gives us one bad reason for this view, which illustrates a common problem in the early days of the study of visual perception. The fact that the image which is projected onto the back of the retina is two-dimensional was taken to show that our immediate visual perceptions must be two dimensional. Berkeley conceives distance as a line which runs from the center of our retina out towards objects, and he notes that there is only one point at which the line hits the retina, no matter how long the line happens to be. This is discussed in Armstrong (1960), Riskin (2002), Morgan (1977).

to estimate the distances of perceived objects, for no one can seriously doubt that we do estimate the distances among objects themselves, and between ourselves and objects, by sight. Berkeley's fellow students of optics had argued that we are able to estimate how far away an object is from our bodies by an innate geometry:

> [I]t is the received opinion that the two *optical axes* (the fancy that we see only with one eye at once being exploded) concurring at the *object*, do there make an *angle*, by means of which, according as it is greater or lesser, the *object* is perceived to be nearer or further off. (IV)

It follows from this account of the ability to estimate distances that a necessary connection holds between an acute angle and a far object and an obtuse angle and a near object. Moreover, Berkeley takes it to imply that experience isn't needed for acquiring knowledge of distances. If there is a necessary connection between the size of the angle where the axes meet the object and the distance the object is from our eyes, then in principle we should be able to work out distances independently of any perceptual experiences.

In contrast to the geometrical account, Berkeley offers an empiricist theory of our knowledge of distances, in which experience provides the cues needed to judge distances. According to Berkeley, the cues we need to judge distance by sight are provided by touch, which he understands broadly to include kinesthesia. We learn to correlate the visual ideas we have with certain sensations that go along with having our eyes in a particular position. When we look at near and far objects, the interval between our pupils changes, and this change is something that we can feel. (Like the natural geometry account, I take it that Berkeley assumes that this process, by which we correlate the sensation, which comes from the position of the eyes, and the visual idea, is (usually) unconscious.)

When he wrote *An Essay Towards a New Theory of Vision*, Berkeley had not yet embraced a full-blown immaterialism.[34] While in 1709 he believed that in vision we immediately perceive only ideas, he had not yet given up the notion that through touch we do come into contact with the physical world. Touch gives us access to the three dimensional world. Touch provides us with the ideas of distance, tangible figure, and solidity, and these ideas allow us to estimate distances:

> Looking at an object, I perceive a certain visible figure and colour, with some degree of faintness and other circumstances, which from what I have formerly observed, determine me to think, that if I advance so many paces or miles, I shall be affected with such and such ideas of touch: so that in truth and strictness of speech, I neither see distance itself, nor anything I take to be at a distance. (XLV)

Berkeley goes on to explain that the connection between visual and tangible ideas is contingent. We learn that the visual idea is frequently associated with the tangible idea in the same way that we learn that the sound of wheels on a road is a sign of a coach. Anticipating Hume's discussion of constant conjunction, Berkeley argues that by constantly observing the visual idea and the tangible idea together, we learn the visual cues that allow us to estimate distances by sight alone. When we judge an object to be a certain distance away from us, say a mile, what we mean by this judgment is that if we were to walk a mile we would be in a position to receive certain tangible ideas (XLIV-XLV).

The constant conjoining of the visual and tangible ideas explains the natural tendency to believe that when we look at and touch a coffee mug, we are seeing and touching one and the same object. According to Berkeley, this is a mistake, a mistake made easy by the fact that we give one and the same names to properties that are really quite distinct. For example, we call both the visual idea of being a cylinder and the tactile idea of being a cylinder "cylindrical." Berkeley's claim that the visual and

[34] I'm following Armstrong (1960) here. See pp. 26-32.

tactile perceptions are of numerically different objects follows from his claim about what kinds of perceptions are available in vision. As we've seen, Berkeley thinks, along with almost everyone else at the time, that the visual ideas are two-dimensional. In fact, Berkeley believes that the only kind of visual input we get is light and colors, arranged in certain ways (CXXIX). From the fact that the visual perception relates us only to two-dimensional objects, while touch relates us to three-dimensional objects, it follows that the objects of vision and the objects of touch are numerically distinct.

Berkeley does not think that the objects of vision and of touch are merely numerically distinct; he also argues that they are specifically distinct, i.e. that the objects of vision and touch belong to different species or kinds: "*The extension, figures, and motions perceived by sight are specifically distinct from the ideas of touch, called by the same names, nor is there any such thing as one idea or kind of idea common to both senses*" (CXXVII, original italics). This claim of Berkeley's can be broken down into two parts: (i) the ideas of sight are different in kind from the ideas of touch and (ii) no ideas, or even kinds of ideas, are shared by vision and touch.

What is the relationship between these two claims? Berkeley seems to think that (i) implies (ii). But it doesn't. (The second claim does imply the first. If there are no ideas held in common by vision and touch, and we believe that there are ideas of both vision and touch, then it seems that these ideas must be of different kinds.) Most people would grant that the ideas of sight and touch differ in their phenomenal character; in this way, the ideas of sight and touch differ in kind. But even if the ideas of the two senses differ qualitatively, it is possible that the two senses have in common an abstract idea. The visual idea of a square and a tactile idea of a square differ in kind (you can have one without having the other), but they still seem to have in common what we might call the geometrical idea of a square. Another possibility is that the visual and tactile ideas differ qualitatively, yet they have in common an idea,

which is the referent of the two ideas. This would not be amenable to Berkeley, since he thinks that the visual idea is a sign of the tangible idea, but it shows that the claim that vision and touch do not have in common any ideas does not follow from the fact that the ideas of the two senses are specifically distinct.

A New Theory of Vision includes arguments against the coherence of the notion of abstract ideas (CXXV), which, if sound, would be reason to accept (ii). I'm not going to reiterate these arguments here, but think it sufficient to point out that few philosophers have found them convincing. In addition to the anti-abstraction argument, Berkeley gives us three different reasons for the second part of the specific distinctness claim; each of which depend upon phenomenological considerations.[35]

First, he appeals to the intuition that a blind man given sight would fail to connect his new visual ideas with his old tangible ones; he would, Berkeley claims, think that the visual ideas were of entirely different properties from his tactile ideas. Second, he offers the following brief argument:

> 1. Light and colors are only perceived by sight; in other words, light and colors cannot be perceived through any other sense, e.g. touch.
> 2. Light and colors are the only immediate objects of sight.
> 3. Thus, there is no idea common to both vision and touch (or any other sense).

Third, he argues that only quantities of the same kind can be added together. For example, we can add two lines together to make a longer line, but we can't add a line to a solid. Neither, Berkeley claims, can we add a visible line to a tangible line, which we recognize when we try and fail to conceive such an addition. Thus, since quantities of like things can be added together, a red line can be added to a blue line for example, and visible and tangible lines cannot be added together, visible and tangible lines cannot be of the same kind.

---

[35] Notice that Berkeley's strategy differs from Locke's, which merely relied on what he assumes to be a common intuition.

For these reasons, Berkeley argues that the ideas of vision and touch are different in kind, and share no common ideas. With these arguments in place, Berkeley turns to the question posed by Molyneux (CXXXII). As we saw above, Locke's answer to Molyneux's question is a rather hasty "no." Locke seems to think that it followed simply from the empiricist principle, which I argued is not the case. Berkeley has a more complete answer. He begins by acknowledging, which Locke did not, that if a square surface perceived by touch and a square surface perceived by vision are of the same kind, then it is possible that the blind man *would* correctly name the objects in front of him, because "it is no more but introduced into his mind, by a new inlet, an idea he has been already well acquainted with" (CXXXIII). If one and the same property is the source of the ideas, then Berkeley seems to be saying, it's possible that the blind man will recognize the new visual idea as coming from the same source as his old tangible idea.

Berkeley realizes that the numerical distinctness of the visual and tangible ideas will not be sufficient for a negative response to Molyneux's question; the ideas must also differ in kind. A philosopher could agree with Berkeley that visual and tactile ideas are numerically different, without believing that the ideas were of different sorts.[36] The idea of a visible line is numerically different from the idea of a tangible line, since you can have one without having the other. But it doesn't follow that these ideas aren't essentially the same kind of idea. If the empiricist is to answer Molyneux with an unequivocal "no," then he must also argue that the visual sphere idea and the tactile sphere idea are of different sorts.

Berkeley points out that if he is right, then Molyneux's question will be unintelligible to the newly sighted man. As we've seen, Berkeley believes that the

[36] What does it mean to be of different sorts? Berkeley is assuming that it is sufficient for the ideas to be of different sorts that they come from/refer to different objects. Whether this is also a necessary condition is not clear.

relationship which holds between the visual and tangible ideas is contingent and acquired by experience; visual ideas represent tangible ideas in the same way that written words represent sounds (CXLIII). If you put a book in the hands of an illiterate man, the words are unintelligible; he has to be taught to match the written words to the sounds he knows. Similarly, when the blind man gains his sight, he must *learn* how to read the visual signs that correspond to the tangible ideas he has.

Conclusion

In Locke's and Berkeley's different responses to Molyneux Question we can trace the evolution of the grounds for the empiricist principle. This development is seen first of all in their different interpretations of the question. Locke was primarily concerned with refuting the nativist, and so he understood the question as one about the source of our ideas. Berkeley, more fully understanding or willing to accept the consequences of a pure empiricist account, recognized that Molyneux's question is really a question about the nature of the perceptual experience, and not just about its origins. Berkeley tries to defend the empiricist principle on phenomenological grounds, by appealing to certain features of what our different sense experiences are like. Second they differ on there understanding of the learning process involved. This is illustrated by the fact that Locke's view appears to be compatible with Leibniz' common sensible response; Berkeley's view is not.

2.3 The Role of Imagination

In the last section of the chapter, I want to examine the role played by the imagination. Perhaps the most famous counterexample to the empiricist principle comes from one of its staunchest defenders, David Hume. One of the most puzzling aspects of Hume's discussion surrounding the missing shade of blue is his nonchalant acceptance of what appears to be a devastating counterexample. I'll argue that there

are good reasons for Hume not to worry about the potential counterexample, reasons which do not apply to the knowledge argument.

But I begin with a discussion of Descartes' theory of the imagination. I discuss Descartes' view in such length because he has the clearest statement of an account of imagination and because his account influenced the empiricists. Another advantage of Descartes' view is that the imagination's role is narrowly defined; Hume's notion of imagination, on the other hand, is much broader, because he subsumes the role played by the intellect in Descartes under imagination.[37] For the purposes of discussing the missing shade of blue—and particularly why Hume thought that imagination could provide the missing shade in the absence of the impression—I think that the differences between Hume and Descartes are not relevant.

2.3.1 The Faculty of Imagination

I'm going to focus on the account of imagination Descartes embraced in his later years, in particular the view expressed in the *Meditations*. (It's worth noting that in earlier writings Descartes conceived of the imagination playing a much larger role in the acquisition of knowledge than the view described here.[38]) The imagination is one of the faculties of the human mind; it is one way to think about things that aren't currently present. In the *Meditations*, the imagination is described as a faculty similar to the senses and the understanding, but distinct from both. Like the senses, and unlike the understanding, the imagination is dependent on the body; like the understanding, and unlike the senses, the imagination is frequently under the subject's control. There are two general kinds of imaginative processes according to Descartes, a passive mode which is modeled on the senses, and an active mode which is modeled

[37] See Garrett (1997) Chapter 1 for a discussion of Hume's notion of imagination.
[38] Dennis Sepper (1996) traces the development of Descartes' theory of imagination, and emphasizes that in his early work Descartes believed that the imagination could help us attain scientific knowledge.

on the understanding. The former mode is illustrated by dreams, in which we passively enjoy a series of images. The latter mode is illustrated by the ability to call up an image of a triangle to help with a geometry problem.

In the following passage, Descartes describes in detail some differences between the imagination and the understanding. In particular, the imagination is limited to operating on images, which are pictorial, while the understanding is able to transcend the limiting features of pictorial representations. Descartes uses his own failed attempts to imagine a chiliagon to argue for a distinction between the faculty of imagination and the faculty of understanding. Since we cannot clearly represent a chiliagon in imagination, but we can clearly represent one (e.g. we can distinguish a chiliagon from other polygons), we must be using a different mental faculty.

> When I imagine a triangle, for example, I do not merely understand that it is a figure bounded by three lines, but at the same time I also see the three lines with my mind's eye as if they were present before me; and this is what I call imagining. But if I want to think of a chiliagon, although I understand that it is a figure consisting of a thousand sides just as well as I understand the triangle to be a three-sided figure, I do not in the same way imagine the thousand sides or see them as if they were present before me. It is true that since I am in the habit of imagining something whenever I think of a corporeal thing, I may construct in my mind a confused representation of some figure; but it is clear that this is not a chiliagon. For it differs in no way from the representation I should form if I were thinking of a myriagon, or any figure with very many sides. Moreover, such a representation is useless for recognizing the properties, which distinguish a chiliagon from other polygons. But suppose I am dealing with a pentagon: I can of course understand the figure of a pentagon, just as I can the figure of a chiliagon, without the help of the imagination; but I can also imagine a pentagon, by applying my mind's eye to its five sides and the area contained within them. (Descartes, *Sixth Meditation*, trans. Cottingham, p. 51)

According to Descartes, he can imagine triangles and pentagons, but he cannot imagine chiliagons. The passage makes fairly clear that the difference is due to his inability to form a picture of a chiliagon for his mind's eye to "look" at. When he

imagines a triangle or a pentagon it is as if he is seeing the figure before him; that is, the content of a mental state when imagining a triangle is the same as the content of a mental state when perceiving a triangle. This is not the case when he tries to imagine a chiliagon: "I do not in the same way imagine the thousand sides or see them as if they were present before me." The content of imagining a chiliagon is different from the content of seeing a chiliagon. The imaginatory content is degraded; all he can call to mind is a "confused representation."[39]

Why can't Descartes imagine a chiliagon? Here is one possible reason.[40] Suppose that in front of him are a chiliagon and a 1001-sided figure. Merely looking at the two figures will not enable him to distinguish the chiliagon from the other figure; the difference between the two figures is too fine-grained.[41] He could, of course, distinguish them on the basis of perception by counting the sides. But this clearly involves more than just perceptual content; it also involves the concepts of numbers up to 1001. This suggests another reason for the failure to imagine a chiliagon. Unlike triangles and pentagons, most humans can't distinguish chiliagons on the basis of perception alone.[42] So in imagining a chiliagon we can't just think about what it's like to *see* a chiliagon, since imagining what it's like to see a chiliagon would not alone distinguish the imagined chiliagon from other polygons. You also need to count the sides. Since the perceptual image alone won't distinguish the chiliagon from the figure with 1001 sides, neither will the imagined image of a chiliagon.[43]

---

[39] That we can have a degraded image of a chiliagon suggests that Descartes held, unlike Berkeley and Hume that pictorial images were not necessarily determinate.
[40] Cf. Tye (1991, pp. 3-4).
[41] The same would be true of an attempt to imagine a triangle with precise measurements, for example, a triangle with sides of 2.15 mm rather than a triangle with sides of 2.14 mm.
[42] I say most humans, since it seems likely that there are some idiots savant who can tell at a glance whether the figure has 1000 or 1001 sides.
[43] As we'll see in the next chapter, some philosophers have argued for an imagination based account of Mary's knowledge of what seeing red is like. If the imagination is modeled on perception, as Descartes

If this – that the perceptions of the chiliagon and the 1001-sided figure are not distinguishable by perception alone – is one of the reasons for Descartes' claim that he cannot imagine a chiliagon in the same way that he can imagine a triangle, then it suggests that Descartes believes that the content of the imagined image is copied or derived from the perceived image.[44] And this account is supported by Descartes' view that the imagination, like the senses, is dependent upon the body. The imagination acts on the images provided by the body: "when the mind understands, it in some way turns towards itself and inspects one of the ideas which are within it; but when it imagines, it turns towards the body and looks at something in the body which conforms to an idea understood by the mind or perceived by the senses" (*Sixth Meditation*, p. 112). An example of an image which conforms to an idea *understood by the mind* which can be imagined is a particular triangle; an example of an image which conforms to an idea *perceived by the senses* is a fantastical creature composed of parts of different animals. As Descartes later says, the contents of many of our imaginings are made up of things that we have perceived in the past.

2.3.2 Hume's Missing Shade of Blue

In his *Treatise of Human Nature,* Hume famously wonders whether a normally sighted person could form an idea of a shade of color she has never seen. Hume imagines a man who had never seen a particular shade of blue; we'll call it "the missing shade of blue." The man is then confronted with a spectrum of very finely discriminated shades of blue from dark to light, with a blank space where the missing shade of blue should be. Hume proposes that, despite never having seen the missing

---

suggests, then this might pose a problem for how we acquire knowledge from imagination (if you think that being able to distinguish the image from similar images is necessary for having knowledge of it).

[44] This account of how the imagined image gets its content is essentially Hume's view, which is discussed below.

shade, the surrounding shades of blue will provide the spectator with the necessary ingredients to imagine what the missing shade of blue looks like:

> Now I ask, whether 'tis possible for him, from his own imagination, to supply this deficiency, and raise up to himself the idea of that particular shade, tho' it had never been convey'd to him by his senses? I believe there are few but will be of the opinion that he can; and this may serve as proof, that the simple ideas are not always deriv'd from the corresponding impressions; tho' the instance is so particular and singular, that 'tis scarce worth our observing, and does not merit that for it alone we shou'd alter our general maxim. (1.1.1.10).

There are two points about this passage which I'd like to call your attention to. The first point is that the man relies on the *imagination* to supply him with an idea of the missing shade of blue. Second, the "general maxim" to which the missing shade of blue appears to be a counterexample is one of Hume's central theses, called the *copy principle*, which states that every simple idea is derived from a simple impression. The question I am most interested in is why, given his general adherence to the copy principle, Hume though that the imagination could so readily supply an idea of the missing shade of blue.[45]

Before discussing this question, let me remind you of the basic elements in Hume's theory of mind. Hume divides perceptions into two sorts, impressions and ideas. Ideas and impressions are distinguished from each other by differing degrees of forcefulness and vivacity; ideas are less forceful and vivacious than impressions. Impressions and ideas come in two basic sorts: simple and complex. Simple ideas are those which "admit no distinction nor separation" (1.1.1.2). By this, Hume means that simple ideas are atomistic in the sense that simple ideas cannot be broken down into distinct parts, but he also thinks that simple ideas bear relations of resemblance to one

[45] The question which most perplexes scholars is why Hume doesn't think this concession affects his arguments against the ideas of cause, substance, personal identity, etc. See Cummins (1978), Fogelin (1984), Russow (1980) for discussion of this point.

another.[46] Complex ideas are made up of simple ideas. To borrow Hume's own example, the idea of the apple is complex since we can distinguish the color, smell, and taste (among other qualities). The color of the apple would be simple, since it cannot be further broken down into parts.

Hume's copy principle is a descendent of the empiricist principle. Here is Hume's statement of the copy principle:

> (CP) All our simple ideas in their first appearance are deriv'd from simple impressions, which are correspondent to them, and which they exactly represent. (1.1.1.7).

The copy principle can be broken down into two parts: (i) the claim that all our simple ideas are derived from simple impressions and (ii) the claim that our simple ideas exactly represent the impressions from which they are derived. Hume later makes clear that the relationship of representation that holds between the ideas and the impressions is one of resemblance.

Assuming, with most commentators, that Hume's distinction between simple and complex ideas is essentially the same as Locke's distinction (i) is a restatement of Locke's empiricist principle.[47] He differs however in how he understands the relation of representation. On Hume's account, unlike Locke's, the relationship of correspondence is one-one; that is, each simple impression gives rise to just one simple idea.

Like Locke, Hume's main style of argument in favor of the copy principle is to ask the reader to come up with a simple idea which lacks a corresponding simple

---

[46] He says, for example: 'Tis evident, that even different simple ideas may have a similarity or resemblance to each other; nor is it necessary, that the point or circumstance of resemblance shou'd be distinct or separable from that in which they differ. *Blue* and *green* are different simple ideas, but are more resembling than *blue* or *scarlet*; tho' their perfect simplicity excludes all possibility of separation or distinction. 'Tis the same with particular sounds, and tastes, and smells. (Appendix)

[47] It's not clear whether Hume wholly accepts Locke's original distinction between simple (ideas/impressions) and complex ideas. Certainly, there are many similarities in how the two understand the relationship between simple (ideas/impressions) and complex ideas. Garrett (1997), however, argues that the two philosophers conceived of the distinction quite differently. See pp. 60-62.

impression. Given the significance of the copy principle for Hume, and the style of argument he uses to support it, his discussion of the missing shade of blue comes as a shock. The thought experiment looks like a straightforward counterexample to the copy principle: the man forms a simple idea of a color of which he has never had an impression.[48]

Hume imagines a situation in which the surrounding shades of blue are similar enough to the missing shade to guide and constrain the man's imagination. But the idea formed in this and similar cases, is not, I want to claim, an idea with new experiential content.[49] It is an idea which refers to the existence of an experience which the subject has not had, but an experience he knows something about. The man knows that there is a discriminable shade which is lighter than the shade to the left of the space and darker than the shade to the right. This is quite a lot of information, and it's enough to rule out a myriad of other possibilities. This explains why the intuition that the man knows what the shade looks like is compelling; if the man was confronted with an impression of the shade, he could fit it in the space like a jigsaw puzzle.

There are two reasons we might find the intuition that the man could form an idea of the missing shade compelling. If concepts of colors were functional concepts, then it would be appropriate to attribute to the man the idea of the missing shade of blue; if ideas of colors are nothing above and beyond the man's abilities to fit the idea into its phenomenological color space, then he has the idea of the missing shade of blue. And perhaps this is what we should say, if we think that ideas are concepts. Not

[48] Personally, I think Hume was too quick to concede the counterexample. I am not confident in my own ability to imagine the missing shade of blue, and there is some evidence that we generally are bad assessors of our imaginative abilities, in particular our ability to generate and manipulate images. See Eric Schweitzgebel (2002).

[49] Many commentators have pointed out that this case may not be so singular. Similar cases could be constructed for other sensations. See Russow (1980), Morreall (1982), and Fogelin (1984) for accounts of the missing shade which discuss the importance of the phenomenological similarities that holds among color shades.

only does the man have a concept that refers to the particular missing shade, he also can distinguish the referent from other possibilities.

There is a second reason for thinking that the man could form the idea of the missing shade of blue, which is more in line with Hume's conception of idea formation. Hume thought the man's idea would come from his formation of an image of the missing shade. By synthesizing the information from the two surrounding shades, the man would be able to form a new idea. Instead of receiving the sense impression in the usual passive way, the man is able to form his own impression using his imagination. And this process of idea formation, unlike the process of concept formation described above, is less threatening to Hume's copy principle. The generation of a functional concept clearly violates the copy principle: there is no sense impression from which the concept can be copied. But in the imaginative process, the man generates a sense impression from which the idea is copied; the only difference from the ordinary case is the origin of the sense impression.[50]

The difference between the concept and the idea is the particular qualitative component. The man can have the concept without having the particular experience of the missing shade of blue, and thus without knowing precisely what enjoying such an experience might be like. And, generally, this is how concepts work. The concept RED does not have a single shade from which it gets its phenomenal content, but is associated with the phenomenology of a range of red shades. When we employ the concept RED we don't need to have a particular phenomenal content in mind. The man does not need to have an image of the missing shade of blue on order to have thoughts either about it, or with it. That is, it is clear that we can refer to things, even

[50] I must acknowledge that there is a major flaw with attributing this proposal to Hume. In the passage quoted, he states fairly clearly that he takes the missing shade of blue to be a counterexample to the claim that for every simple idea there is a corresponding impression. In my proposal, the man generates his own simple impression, and thus the thought experiment is not a genuine counterexample. See Garrett (1997, 50-52) for a suggestion which is similar to mine.

phenomenal things, without knowing much about them.[51] The man can refer to the missing shade of blue, even if he has no idea associated with it. He can think of it under the concept THE MISSING SHADE OF BLUE. But having the idea requires having the relevant experience.

Conclusion

Hume's discussion of the missing shade of blue is designed as a counterexample to the intuition that all ideas come from corresponding impressions. The counterexample gets its plausibility from the fact that the boundaries of the missing idea are provided by the ideas the man does possess. The ideas of the lighter and darker shades the man has provide fairly precise instructions for what he should imagine. In so imagining, the man generates his own impression of the missing shade, which gives rise to the idea. There is still a one to one correspondence between the sense impression and the idea, and so the core of the copy principle, that there must be a sense impression for every idea, is safe from this putative counterexample.

Summary of Chapter Two

In this chapter we've looked at three thought experiments proposed during the modern period. These three thought experiments were chosen because each highlights a different aspect of the empiricist principle: the claim that all our knowledge of the external world comes from sense experience. The empiricist principle is closely tied to what I called in the first chapter *the central intuition* of the knowledge argument, the intuition that Mary must have the experience of seeing color, before she can know what seeing color is like. The purpose of this chapter is to examine the motivations that led the empiricists to embrace the empiricist principle, in the hope that this might

[51] This is discussed in much more detail in the beginning of Chapter Four.

lead to a suggestion for why so many people feel the pull of the central intuition so strongly.

In the first section, we saw that Locke's support for the empiricist principle is in large part due to his frustration with the nativist position that was dominant at the time. He argued that the nativist lacked any empirical support for his position, and he argued that on any reasonable account of innate ideas they are explanatorily inert. But Locke also had some positive support for the empiricist principle. Locke's own thought experiments are similar to two influential thought experiments in contemporary philosophy, Nagel's bat and Jackson's knowledge argument. Unfortunately, Locke's two thought experiments don't go any further than pumping intuitions. His thought experiments provide intuitive support for the empiricist thesis, but they don't go any way towards explaining it.

This leads us to Berkeley's discussion of Molyneux's Question. There was some reason to think that Berkeley's discussion of Molyneux would give us some insight into why experience is necessary for knowledge, because Berkeley offers a sophisticated explanation for why Molyneux's newly sighted blind man can't identify a cube or a sphere by vision alone. Berkeley's response to Molyneux is especially interesting for my purposes, since it depends crucially on a claim about the nature of perceptual experience. But Berkeley is primarily interested in (i) establishing a difference in kind between visual and tactile experience and (ii) showing that the relationship between the two experiences is contingent and acquired by us through repeated experiences. While his suggestions are interesting, it is not clear that they are directly relevant in explaining why Mary can't know what colors are like, since this is not a cross modal case. However, as we'll see in Chapter Four, my own proposal for the central intuition is Berkeleyan in spirit.

Finally, we examined the role of imagination in acquiring knowledge, and Hume's own counterexample to his version of the empiricist principle. The missing shade of blue thought experiment is crucially different from the thought experiment behind the knowledge argument. I argued that it seems plausible that the man could generate for himself a sense impression which would correspond to the idea of the missing shade of blue. But, with rare exceptions, most philosophers don't think that Mary would be able to generate her own experience of color. None of Mary's experiences while she is in the black and white room are similar in the right sort of ways to the experiences she'll have when she leaves the room. She has no experiences that can help guide or constrain her attempts to imagine what seeing colors is like. Imagination is not powerful enough to help her.

Chapter Three
Knowing How and the Explanatory Constraint

Introduction

In the last chapter, we explored the historical roots of the empiricist claim that knowledge requires experience. We saw that the empiricists themselves did not have a satisfying explanation of why experience is necessary for knowledge. In this chapter, we'll look at a response to the knowledge argument that directly addresses this issue. This response to the knowledge argument is inspired by the work of Gilbert Ryle. In *The Concept of Mind*, Ryle argues for a category of knowledge distinct from propositional knowledge, which he calls "know-how." The distinction between propositional knowledge and know how has been used by a number of philosophers in responding to the knowledge argument. In this chapter, I examine the strategy employed by these philosophers. I argue that a significant, and undervalued, benefit of this response is that it offers an explanation for why Mary cannot have knowledge of what it's like to see red before having the experience of seeing red. Nonetheless, I will contend that the know-how response fails due to its inability to accommodate a set of counterexamples.

The know how response is appealing to me because it offers an explanation of the *central intuition*: the intuition that Mary learns something new when she leaves her room. As we saw in the first chapter, most people find the central intuition appealing and try to accommodate it. Daniel Dennett and Paul Churchland, however, deny the central intuition, and I take this denial as a genuine challenge to those who want to

accommodate the central intuition.[1] Dennett and Churchland argue that we, as humans with average cognitive powers and limited knowledge of visual science, have no idea what it would be like to be in Mary's position, with her awesome cognitive powers and complete knowledge of visual science.

As Daniel Dennett says:

> That [Mary would learn something when she leaves her room] is how almost everyone imagines this thought experiment—not just the uninitiated, but the shrewdest, most battle-hardened philosophers. Only Paul Churchland has offered any serious resistance to the *imag*e, so vividly conjured up by the thought experiment, of Mary's dramatic discovery. The image is wrong; if that is the way you imagine the case, you are simply not following directions! The reason no one follows directions is because what they ask you to do is so preposterously immense, you can't even try. The crucial premise is "She has all the physical information." That is not readily imaginable, so no one bothers. (*TSAM*, 60)

Dennett claims that if we really think about what it means to know *all* the physical truths-- not just the truths which are currently known, but all the physical truths there are regarding color vision-- we'll realize that we can't do it. No human has ever been in a situation remotely like Mary's, and it's hard to know where to begin in trying to imagine her situation. Two problems make imagining what it would be like to be in Mary's situation especially difficult: (i) we don't know what a completed science of color vision would be like and (ii) we don't know what kind of epistemic position a subject who has complete knowledge of color vision would be in.

As for (i), it's not clear how we could tell when a science counts as complete. We currently haven't completed any scientific bodies of knowledge. And the science of color vision will be particularly complicated, since it will include many different fields, e.g. neurobiology, psychology, and physics. Presumably, a complete science of

[1] See Dennett (1991), (2007) and Churchland (1989). Dennett (1991) in excerpted in *TSAM* and Churchland (1989) is reprinted with a postscript.

color vision will have integrated these various disciplines. But whether the integration will take the form of a reduction, or whether the disciplines will be related in some other way, is yet to be determined.

Even if we have a good sense of what a completed color science will be like, this still won't help with (ii). Even with a conception of a completed theory of color vision, it's very difficult to conceive what it would be like to have all that knowledge. In particular, it is very difficult to be certain about whether Mary would or would not be able to deduce what it's like to see red from what she knows while still in her room.

That is, I'm sympathetic to Dennett and Churchland's claim that we, as humans with ordinary mental powers, aren't good at imagining omniscient Mary's situation. In addition to the difficulties in properly conceiving Jackson's thought experiment, I think that in general it's good methodology to offer explanatory grounds for one's intuitions. And finding explanatory grounds for the central intuition is doubly useful. First, it will get us past the stalemate between those who accept that Mary learns something new and those who deny it.[2] Second, providing some independent support for the central intuition could help the physicalist respond to the dualist, who explains Mary's new knowledge by appealing to her contact with nonphysical properties. Ideally, the physicalist would both block the knowledge argument and offer the dualist a compelling *reason* for why Mary can't know everything about color vision despite her complete physical knowledge that doesn't rely on an antecedent commitment to physicalism.[3]

[2] Of course, Dennett and Churchland and those who accept the central intuition are ultimately on the same side; they are all trying to defend physicalism. So the stalemate doesn't affect the soundness of the knowledge argument; if either side of the stalemate is correct, then the argument is unsound. But, as I hope has become clear, I am interested in issues underlying the knowledge argument, and not merely with its challenge to physicalism.

[3] I offer my own proposal in the next chapter.

On these grounds, I take the possibility Dennett and Churchland raise, that Mary could work out what it would be like to experience colors from the physical and functional information she possesses, to place a constraint on a satisfying solution to the knowledge argument. It is not enough for the physicalist to accommodate the central intuition; she also needs to explain *why* Mary cannot know what it's like to see red before leaving the room. I will call this condition on a response to the knowledge argument the *explanatory constraint*.

In the next section, I explain the know how account, and how it meets the explanatory constraint. In section 3, I examine each of the four claims made by the proponent of the know how account. I defend the first three claims, but argue that the know how account ultimately fails because it's possible for Mary to have knowledge of what an experience is like without having any know how. In other words, the know how account provides sufficient conditions for knowing what an experience is like, but not necessary ones. In the fourth and final section, I look at an alternative to the know how account, and argue that it fails the explanatory constraint.

3.1 Knowing How and the Knowledge Argument

3.1.1 Ryle's Distinction

To understand Ryle's insistence on a distinction between knowing how and propositional knowledge it helps to keep in mind the goal of *The Concept of Mind*. Ryle conceives of his project as a rebuke against the prevailing orthodoxy, which in the chapter titled "Knowing How and Knowing That" he calls "the intellectualist legend."[4] Ryle is not as precise as he could be in his description of the intellectualist legend, but a couple themes are clear. First, the intellectualist legend says all

[4] See Snowdon (2004) for an abbreviated discussion of the intellectualist legend. It is unlikely that any philosopher actually subscribed to the all the doctrines included in the legend.

knowledge is fundamentally propositional knowledge.[5] Second, on the intellectualist legend acts are conceived on a two-process model. Each act that is intelligently performed is paired with a corresponding mental process. Presumably, it is because each act is paired with a mental process that all knowledge is fundamentally propositional (1949, 29-34).

Ryle is concerned with activities that can be described by what he calls the "intelligence epithets," words like "intelligent," "silly," "stupid," "prudent," or "shrewd." (For ease of exposition I will call these activities "intelligent acts," by which I mean that their execution can be described by one of the intelligence epithets.) The intellectualist legend takes theorizing as the model for mental activity, especially mathematical and scientific theorizing. The aim of these subjects is "the knowledge of true propositions or facts," and the intellectualist legend assumes all intelligent activities share this aim (1949, 26). According to the intellectualist legend, propositional knowledge is what distinguishes human minds from animal minds, and human activities are to be classified as mental activities only if they are guided by knowledge of propositions. On this view of mental acts, even non-theoretical activities, like cooking or playing the fiddle, are made possible by the actor's grasp of true propositions. So, for example, my knowing how to cook a roast consists in my knowledge of certain propositions, for example: that the meat should be tied in an even shape, that it should be first seared in oil, that the oven should be set at 350 degrees, etc. In what follows, I'll call the view that all knowledge is fundamentally propositional the *propositionalist* account.

In addition to the claim that all knowledge is propositional, the intellectual legend is committed to the claim that each intelligent act embodies two processes.

[5] A terminological note: I use "knowledge" as a neutral way to refer to both propositional knowledge and know how.

Ryle seems to think that trying to model the knowledge demonstrated by cooking and fiddle-playing on theoretical knowledge requires distinguishing between the acts and the thoughts that precede them. When we act intelligently, we first reflect on what we are about to do and how we should best go about it, and then we execute the action. Before I frost the cake, I take some time to think about what I am about to do. I consider many propositions, e.g. that the cake is sufficiently cool, that the frosting will cover the whole cake, that there enough room on the counter, etc. and then I frost the cake. It is by taking into considerations these (or similar) propositions that the act of frosting is intelligently performed (1949, 29).

Ryle concedes that we often reflect on our actions before we act, but he argues that the antecedent reflection isn't always necessary in order to act intelligently: "the absurd assumption made by the intellectualist legend is this, that a performance of any sort inherits all its intelligence from some anterior internal operation of planning what to do" (1949, 31).[6] The absurdity of assuming that all intelligent performances require a preceding mental plan leads Ryle to propose another distinct category of knowledge, *knowing how*. Knowing how does not consist in bearing a relation to a proposition, but in having certain abilities: "When a person is described by one or other of the intelligence-epithets such as 'shrewd or 'silly' or 'prudent' or 'imprudent,' the description imputes to him not the knowledge, or ignorance, of this or that truth, but the ability, or inability, to do certain things" (1949, 27). If I say that Sally is a shrewd bridge player, I ascribe to Sally an ability to frequently win her bridge hands, and not knowledge of any particular truths.

But not just any ability will count as a kind of knowing how. Because I am bigger and stronger, I have the ability to beat my six year old nephew at arm wrestling,

[6] Nor would it be sufficient. We can all recall times that despite our meticulous planning of an act, the execution goes awry.

but this mere ability doesn't count as an example of knowing. It doesn't count as an example of knowledge because I can have the ability to beat my nephew at arm wrestling even if I don't know what arm wrestling is. Ryle's own examples of know how, which include knowing how to apply grammatical rules, tie reef-knots, fish, play chess, prat-fall, multiply and tell a joke, at first seem disparate. Fishing and prat-falling crucially involve mastery over one's body. Grammar and multiplication primarily involve the application of rules. Telling a joke is a social act. Knowing how to tie a reef knot is relatively straightforward, compared to knowing how to fish. But what these abilities have in common, and which my ability to beat my nephew at arm wrestling lacks, is that each of Ryle's examples can be done more or less intelligently.

Knowing how is a certain kind of ability, and a category that includes only those abilities that can be executed more or less intelligently. The intelligibility condition rules out abilities like lifting fifty pounds.[7] Knowing how is characterized by two features. First, the way we typically acquire know-how is different from how we typically acquire propositional knowledge. Tying reef-knots, fishing, playing chess, prat-falling, multiplying and telling jokes are all typically acquired by practice. You can be told the correct way to cast a fly, but it will take practice to know how to do it; as Ryle says, "we learn *how* by practice, schooled indeed by criticism and example, but often quite unaided by any lessons in the theory" (1949, 41).

Of course, we often do get lessons in the theory behind the action, but these lessons aren't necessary, according to Ryle. That is, knowing how does not reduce to propositional knowledge. Ryle illustrates this point with a comparison of two ways a boy can learn how to play chess. In the first case, the boy is explicitly instructed in all the rules of the game. He is told how each of the pieces move, opening strategies, that

[7] Although even examples of brute strength don't fall neatly into the unintelligent category. Knowing the proper way to lift heavy objects (bend the knees, not the back) is an intelligent way to lift fifty pounds.

the goal is to checkmate the king. In the second case, Ryle imagines a boy who picks up the rules from watching and playing the game himself. The boy is given no explicit instruction, but learns how to play through trial and error.

How the boy acquires the ability to play chess differs in these two cases. But in both of these cases, Ryle claims, it is possible that the boy knows how to play chess without being able to cite the rules. In the first case, the boy will probably need to begin by reciting the rules to himself before he moves a piece, but eventually the rules will be internalized. The boy will no longer have to think to himself "the bishop move diagonally" before deciding whether or not to move the bishop; he'll just move the bishop diagonally. He could, Ryle claims, forget what the rules are. And in the second case, the boy never explicitly learns the rules. So in both cases, the boy knows how to play chess despite his inability to explain the rules that guide his behavior. Ascribing the know how requires only that the boy "normally does make the permitted moves, avoid the forbidden moves and protest if his opponent makes forbidden moves…So long as he can observe the rules, we do not care if he cannot also formulate them" (1949, 41). The attribution of knowing how is not justified by the content of a subject's mental state, but is justified by what the subject is able to do.

In describing the special features of knowing how I've stuck with Ryle's own example, but it is a poor one. First, there is a problem with Ryle's story of the acquisition. It's hard to imagine how the boy in the second case picks up the rules without explicitly representing them to himself. The most plausible way of making sense of the patterns of movement of the pieces would be to think of general rules that apply to each of the different shapes. If this is true, then the acquisition would essentially involve propositional knowledge. Second, there is a problem with the claim that it's possible that the boy is unable to explain the rules of the game. It's hard to imagine a situation in which the boy could not cite the rules behind the movements

of the pieces, because the rules are so simple. (It's not difficult to imagine a case in which the boy employs some impressive strategy to check the king, but is unable to explain why he chose to do what he did; the boy intuits the moves to make. But knowledge by intuition is not the type of knowing that knowing how is supposed to account for.) Both special features of knowing how, the acquisition of the ability to play chess and the inexpressibility of the knowledge behind it, are implausible in this situation. Because of this, chess playing isn't a clear example of having know how without having propositional knowledge.

But there are much better cases. Knowing how to ride a bike is only acquired by getting on a bike.[8] No matter how much explicit instruction you receive, you won't learn how to ride a bike without getting on the bike and practicing. The attribution of knowing how to ride a bike is justified by one's ability to ride a bike, not by one's knowledge of a description of how to ride a bike. Someone might be able to elaborately describe the sorts of actions that occur during a bike ride, but if they cannot themselves ride a bike then they cannot be ascribed know how. A more complicated example is knowing how to cook. Cooking is an exercise that can be done more or less intelligently, and being able to cook intelligently requires lots of practice. One acquires a set of cooking skills by practicing; it would be difficult if not impossible to acquire these skills from reading cookbooks. And cooking know how needs to be demonstrated to be justified; reciting recipes wouldn't show that one knows how to cook. On a more pragmatic level, we don't hire lifeguards on the basis of a written test alone; we make sure they can pass a swimming test.[9]

---

[8] I'm assuming that in the paradigmatic cases of know how this is a contingent fact about how humans learn how to do things. It seems possible that differently structured beings could learn to ride a bike without practice. The issue of whether our ways of knowing what the experience of seeing red is like is likewise a contingent fact about how we know things is discussed in more detail in Chapter Four.

[9] Thanks to Tamar Szabó Gendler for this last example.

These cases seem to me to be examples where the ability is both necessary and sufficient for having know how. And they have both the special features that Ryle thinks characterize know how. Their acquisition comes from experience, and the knowledge manifested by the abilities cannot be completely expressed in words.

3.1.2 The Know How Account

Ryle's distinction between knowing how and propositional knowledge is the inspiration for an influential response to the knowledge argument, which I'll call *the know how account*, that has been offered with slight variations by David Lewis, D.H. Mellor, and Laurence Nemirow.[10] The strategy behind the know how account is fairly simple. Mary learns something when she leaves her room. But she doesn't gain any new propositional knowledge. What Mary acquires are the abilities to recognize, recall, and imagine colors. These abilities constitute know how. According to the know how account, the knowledge argument is unsound. The second premise—Before leaving her room, there is a truth Mary doesn't know—is false. Mary knows all the truths about colors; all Mary lacks are abilities. In the next section, the know how response will be examined in detail. But first I'd like to show how this strategy meets the explanatory constraint, which requires that responses to the knowledge argument explain why Mary can't know what it's like to see red before she leaves her room.

The know how account is attractive to physicalists and nonphysicalists alike. Mellor endorses the know how strategy because it offers some explanation for the central intuition. In the following passage, he argues that it is the *only* good explanation for why Mary must leave her room to learn what red looks like:

> The inability to say what our experiences are like would be a complete mystery if knowing what they are like were knowing facts, i.e. knowing

[10] See Lewis (1988), Nemirow (1990), Mellor (1992). Lewis (1988) is reprinted in *TSAM*.

> that certain propositions are true. For why, if it is, can we not express these propositions as we can express others, including other propositions about our experiences? For example, I can easily say what I know when I feel warm or taste something sweet: I can say 'I feel warm' or 'This tastes sweet'. Those facts about my experiences are as easily stated as facts about anything else. Sometimes admittedly we know facts we cannot state because we lack the right words. But lack of words is not the problem here, and finding or inventing words is not the solution. It is not my expressive inadequacy, or that of English, which stops me saying what I know when I know, not only that I feel warm, but what feeling warm is like. (1992, 8)

Mellor points out that while we can make certain claims about our experiences, there are certain features of our experiences that we cannot describe. He distinguishes claims like "I feel warm" and "This tastes sweet" from claims about what our experiences feel or taste like. The former claims express propositions about what properties our experiences have, but they don't express what having an experience with the properties of being warm or tasting sweet are like. Unlike the former, these latter claims are not linguistically expressible; that is, they cannot be expressed by sentences.

Mellor's argument can be stated more formally as follows:

> (1) If knowing what an experiences is like is propositional knowledge, then we would be able to express this knowledge linguistically.
> (2) We cannot express knowledge of what an experience is like linguistically.
> (3) Therefore, knowing what an experience is like is not propositional knowledge.
> (4) Knowing what an experience is like is either propositional knowledge or know how.
> (5) Knowing what an experience is like is know how.

I agree, of course, with premise (2); if we could tell Mary what seeing red is like then there would be no knowledge argument. And for present purposes, I'm willing to grant premise (4). The problem with Mellor's argument is the first premise, which relies on the assumption that all propositional knowledge is linguistically expressible. While there is much disagreement about the nature of propositions, the following

features seem uncontroversial. Propositions are the objects of propositional attitudes. If I believe that Ithaca is gorgeous, then I bear a relation of belief to the proposition that Ithaca is gorgeous. If I hope that the Kansas City Chiefs will make it to the play-offs, then I bear a relation of hope to the proposition that the Chiefs will make it to the play-offs. Propositions are shareable. When you believe that Ithaca is gorgeous, you and I believe the same proposition. If you also hope the Chiefs will make it to the play-offs, we both hope that the same proposition will come true. Propositions are the same across different languages: when Pierre believes that Ithaca est jolie, he believes the same proposition as you and I. Finally, propositions are the bearers of truth values. Sentences are true or false in virtue of the propositions they express. My belief that Ithaca is gorgeous is true because the proposition that Ithaca is gorgeous is true.

For these reasons, propositions are sometimes identified with the meaning of sentences. And if sentence meaning and propositional content are identical, then Mellor's assumption that all propositional knowledge is linguistically expressible would follow. For all propositions would be tied to sentence meaning. But there are reasons for thinking that propositional content should not be so closely linked to linguistic content. Even if it's true that the meaning expressed by sentences is propositional, it doesn't follow that all propositions are expressed by sentences. There could be propositions that can't be expressed linguistically. That is, there is nothing about the uncontroversial features of propositions—that they are shareable and are the bearers of truth values—which preclude them from being linguistically inexpressible. On this more expansive view of propositions, the propositions expressed by sentences would be a subset of the set of all propositions. [11]

[11] Mellor also assumes that all facts are linguistically expressible. I'm not sure why one should think this. It's at least possible that there are facts about the world that are beyond human conceptual abilities to grasp, and such facts would not be expressible in any human language.

Despite the failure of Mellor's argument, there is still a good case to be made on behalf of the know how account. Instead of looking for a deductive argument, we should take the account to be a form of inference to the best explanation for the central intuition. There might be other alternatives, but the know-how account explains both why Mary must have the experience herself to complete her knowledge, and also why she cannot be told what it's like to see colors. Mary must experience colors herself because this knowledge consists in abilities and we don't acquire abilities in the same way we acquire propositional knowledge. And she cannot be told what experiencing color is like, because the knowledge that comes with having abilities is ineffable. You can read an entire manual on how to ride a bike, but you aren't going to learn how to ride a bike without having the experience of riding a bike.

Before we look at problems for the know-how account, I want to reiterate its explanatory power. The know how account can explain why Mary can't know what it's like to see red while in her room despite all her scientific knowledge: her new knowledge is of a different sort. But this new kind of knowledge is not mysterious at all; it's like knowing how to ride a bike or filet a fish. No one argues that knowing how to filet a fish shows dualism is true. By relying on these ordinary kinds of cases, the know how account provides a model for how to understand knowing what our experiences are like. This model has two salient features. First, the acquisition of know how is typically different from how we acquire propositional knowledge. Second, the mode of acquisition is special because of its resistance to linguistic expression. Together, these two features explain why Mary cannot be told what seeing red is like, and why she must leave her room to find out.

3.2 Defending the Know How Account

Defending the know how account requires defending the following four claims. (KH1) The distinction between knowing how and propositional knowledge is a genuine distinction between kinds of knowledge. (KH2) Mary gains the abilities to recognize, imagine, and remember colors when she leaves her room. (KH3) The abilities to recognize, imagine, and remember constitute know how. (KH4) The know how constituted by these abilities is identical to knowing what it is like to see colors. In what follows, I discuss each of these claims in detail; for ease of exposition, I group (KH2) and (KH3) and call the two claims, "the ability thesis."

3.2.1 Claim (KH1): The Distinction between Knowing How and Propositional Knowledge

Without a category distinction between know how and propositional knowledge the know how account does not succeed as a response to the knowledge argument. The know how account denies premise 2 of the knowledge argument—Before leaving her room, there is a truth Mary doesn't know—by denying that what Mary learns when she see red for the first time is a truth. But if know how is reducible to propositional knowledge, then this strategy won't work. In gaining know how, Mary will gain propositional knowledge. This is problematic for two reasons. First, Lewis and Nemirow assume that if Mary gains propositional knowledge, then the knowledge argument succeeds and physicalism is false.[12] Second, without a distinct category of know how the account loses its explanatory appeal. The know how account explains why, despite all her scientific and descriptive knowledge, Mary can't know what colors look like while in her room by appealing to a category of knowledge, which is different in kind from the scientific and descriptive knowledge.

[12] As we've seen in Chapter One, not all physicalists agree with this assumption. At the end of this chapter I discuss a version of the know how account which does not depend on a distinction between propositional knowledge and know how.

Ryle argues for an independent epistemic category of know how by claiming that such a category is necessary to stop a vicious regress:

> The crucial objection to the intellectualist legend is this. The consideration of propositions is itself an operation the execution of which can be more or less intelligent, less or more stupid. But if, for any operation to be intelligently executed, a prior theoretical operation had first to be performed and performed intelligently, it would be a logical impossibility for anyone to ever break out of the circle. (1949, 30)

Ryle's argument seems to be roughly this.[13] According to the intellectualist legend, when an agent performs an act intelligently, she contemplates a proposition before performing the act. But contemplating a proposition is itself an act which can be performed more or less intelligently. Thus, this act must itself be preceded by an act of contemplating a proposition, and a vicious regress begins.

The important, and contentious, background assumption is that the propositionalist account of knowing how assumes that intelligent acts are always preceded by acts of contemplation. As Carl Ginet has pointed out, it is manifestly false that all our intelligent acts are accompanied by an act of contemplating the appropriate proposition.[14] To borrow Ginet's example, someone can exercise her propositional knowledge *that she can get the door open by turning the knob and pushing*, without explicitly formulating the proposition. But, Ginet argues, the fact that we sometimes perform actions without first contemplating what we are about to

[13] See Snowdon (2004) and Stanley and Williamson (2001) for critical discussion of the regress argument.

[14] See Ginet (1975), pp. 5-9. The regress argument appears redundant. Ryle agrees with Ginet that people can perform intelligent actions without contemplating the associated propositions, and he uses these examples to his advantage. The paragraphs leading up to the "crucial objection" argue by example for a class of intelligently performed acts that are not preceded by the entertainment of a proposition. He says, for example, that it's "possible for people intelligently to perform some sort of operations when they are not yet able to consider any propositions enjoining how they should be performed." If it follows from the propositionalist account of knowing how that all intelligent actions are preceded by the contemplation of a proposition, then this alone would be reason to reject the propositionalist account.

do doesn't show that the propositionalist account is wrong. This kind of example shows that there are acts which are the result of one process not two. The view that knowledge is propositional is independent from the claim that an act is intelligent if and only if it is preceded by the contemplation of a proposition. Ryle is correct to reject the intellectualist legend and its commitment to a two process explanation of intentional action. But the two process explanation has no bearing on the propositionalist account. As Ginet suggests, the propositionalist can explain an agent's actions as a "manifestation" of propositional knowledge (1975, p. 6).

Ryle needs a further argument that the propositionalist is committed to a two process theory of action. The regress argument's main purpose is to show that an independent category of knowledge is necessary to block the regress. But the regress argument relies on a theory of action which no one should accept. In the rest of this section, I will offer a defense of the distinction which does not rely on a dubious theory of action.

In order to defend a distinction between know how and propositional knowledge, I need to address some counterexamples that Paul Snowdon (2004) advances against the claim that know how consists in having certain abilities. As we've seen, the difference between knowing how and propositional knowledge is that the abilities are supposed to be necessary and sufficient for having know how. Snowdon offers five examples of situations in which someone lacks the ability to do x, but still knows how to x.[15] In four of these cases, Snowdon explicitly states that although the subject cannot himself perform x, he can tell someone else how to do x. For example, Snowdon imagines that the world's greatest chef is in a horrible car accident and can no longer make his excellent omelets. But, Snowdon contends, the world's greatest chef still knows how to make an excellent omelet and "if you wish to

[15] See also Ginet (1975), p. 8.

learn how to make an omelette you should consult [him]" (2004, p.8). He imagines another case in which a maid, having grown up in the royal household, knows how to properly address the queen. However, because of nerves, every time the maid is around the queen she develops a stutter which prevents her from uttering the proper address. Even though the maid cannot herself properly address the queen, contends Snowdon, she still knows how to address the queen.

I have two points to make about this type of counterexample. First, I don't share Snowdon's intuitions that in each of these cases the subject obviously has know how. On the one hand, I admit that it's not clearly wrong to ascribe the chef knowledge of how to make an omelet, because he certainly has a lot of propositional knowledge about omelet making. And there are his past demonstrations of making omelets, which justified his pre-injury claim to know how. But on the other hand, it seems to me wrong to say that the chef knows how to make an omelet. He knows a lot *about* making omelets, but this is different from knowing how to make an omelet. The chef can't be ascribed know how, because he can't do it.[16] We don't ascribe know how to people who haven't demonstrated the ability, and we are rightly hesitant to ascribe know how when the ability hasn't been exercised in a long time. The justification for being correctly ascribed know how is in the exercise of the activity. Since the chef cannot exercise his ability, there is a sense in which he cannot be ascribed know how.

One of the features of know how is that it's possible to have it without having the associated propositional knowledge. In some cases, our acts are themselves

[16] I recognize that this is a contentious claim. I think resistance to the idea that the chef lacks know how is due to philosopher's bias towards propositional knowledge. The chef (and maid) have propositional knowledge and this seems to many people to be sufficient for know how. And, as Tamar Szabó Gendler has pointed out, there are some difficult cases, which I'm ignoring. So, for example, what to say about the chef who merely breaks his arm? Does he stop knowing how to make omelets during the time his arm is in the cast? I'm not sure what to say about such cases, but I also doubt that we will ever have necessary and sufficient conditions for the attribution of knowledge.

justification for the attribution of knowledge. There are musicians who converse very intelligently about the practice of their craft, and there are those who just play. In both kinds of cases, the musician is correctly ascribed know how. Ryle takes this fact, that people can perform intelligent activities without being able to describe what they were doing, as evidence that no propositions are involved. This, perhaps, is a bad inference, and is discussed in detail below. But Snowdon's counterexamples all lack this special feature of knowing how. In each of Snowdon's cases, we are justified in attributing the know how because the subject can verbally describe how to do x. But the case which interested Ryle was one in which the subject lacks the ability to explain the theory behind her activity.

This isn't to say that Ryle thinks that only cases where the subject is unable to express her knowledge counts as know how. I take it that Ryle wants to treat cases in which the agent can express her knowledge like the ones in which she cannot. Suppose there are two musicians, but only one can describe why and how she does what she does. Both are cases of know how, since in both cases the musicians have the abilities. The fact that there are cases where the subject has abilities without being able to describe them, combined with the assumption that propositional knowledge is expressible, is supposed to show that a subject can have know how without having propositional knowledge. This brings me to my second point: all Snowdon's cases show is that a person who loses an ability can retain propositional knowledge; they don't show that the know how can be retained after losing the ability. This explains the sense in which we are willing to ascribe know how to the armless chef: we're really just ascribing the propositional knowledge associated with the ability.

In addition to arguing against the necessity of having the abilities, Snowdon offers four counterexamples that allegedly show that having the ability to x is not

sufficient for knowing how to x. One of these counterexamples is a feat of strength.[17] Generally, we don't consider feats of strength—like the ability to do fifty push-ups or lift three hundred pounds—to be examples of know how. We say that someone can do fifty push-ups, but it sounds odd to say that someone knows how to do fifty push-ups. This case is somewhat similar to one discussed by Stanley and Williamson (2001, 414-15). Stanley and Williamson also point out that not every action counts as know how; we don't, for example, think that the fact that we can digest food entails that we know how to digest food. Stanley and Williamson conclude from this example that we should restrict the range of actions to intentional actions, a restriction which would also rule out feats of strength (2001, 415).

Again, I don't share Snowdon's confidence about what should count as know how. It might be a bit awkward, but it isn't obviously wrong to attribute knowing how to do fifty push-ups to someone who has the ability. Some people don't know how to do fifty push-ups because they lack the strength. But others don't know how to do fifty push-ups, despite having the strength, because they don't know what a push-up is. If it's natural to say that some people don't know how to do a push-up, then it seems that there must *be* know how they are missing. (Knowing what a push-up is arguably counts as a piece of propositional knowledge. But the point here is just that this is not a case in which the ability is there, but the know how is missing.)

But simpler cases, like the ability to lift three hundred pounds, would serve Snowdon's point. So I agree that mere ability will be insufficient as a condition on know how, and some more sophisticated notion of ability is needed. Ryle's own limit, that we consider only those actions that can be described by one of the intelligence epithets, would also rule out feats of strength. And, as Stanley and Williamson seem to think, Ryle's limit to intelligent acts might correspond to their own restriction to

[17] Ginet offers a similar example, (1975), p. 8.

intentional actions. (This shifts the debate to delineating the category of intentional actions, which is also controversial. For example, Stanley and Williamson think that winning the lottery is an unintentional act, at least in part because it cannot be intelligently performed (2001, 415). But, it seems to me, a lottery winner does know how to win the lottery. She also knows how to lose the lottery. The relevant event here is not the winning (or the losing), but the entering. Knowing how to win the lottery entails that one knows how to play the lottery. And, like doing push-ups, this is a kind of knowing that not everyone has.[18]

A full defense of knowing how requires some principled distinction between mere abilities and abilities which are candidates for know how. Not only do we want to rule out actions like lifting three hundred pounds or digesting a meal, we also want to distinguish between instances of simple agent causation and intentional actions.[19] We want, for example, to distinguish the case of a novice dart player who hits a bulls-eye through luck from the player who consistently hits the bulls-eye. Finding and defending such a distinction is beyond the scope of this dissertation, in part because such a defense runs into the same problems that face any account of knowledge. Ascriptions of knowledge are highly contextual, and ascriptions of knowing how are no exception. (This sensitivity to context could explain, for example, why in certain instances it sounds wrong to attribute know how to someone who can do fifty push-ups, and in other instances it sounds natural.)

The other three examples in which having the ability to do x is insufficient for knowing how to do x are cases in which someone has a general ability, but has not yet encountered the specific situation in which the ability will be exercised. So, for example, Snowdon imagines a person in an unfamiliar room which he hasn't yet

[18] Cf. Noe (2004) pp. 120-22. See also his (2005).
[19] See David Carr (1979) for a defense of knowing how that argues for such a distinction.

explored and so does not know how to get out of. (Assume that the door by which he entered locked behind him.) At the end of the room around the corner is an unblocked door, which he can easily open. So, he can escape the room, but he doesn't yet know how to escape the room. In a similar example, Snowdon imagines sight-reading an unfamiliar piece of music. Before playing the piece, Snowdon has the ability to sight-read the music, but until he plays the piece it seems wrong to claim he knows how to sight-read that particular piece of music.

These cases bring up an interesting issue: how are general abilities to do x, related to particular instances of doing x? In the above cases, it seems perfectly clear that the man knows how to get out of rooms, and that Snowdon knows how to sight-read music. Presumably, the man and Snowdon have frequently demonstrated their respective abilities: the man has left many rooms; Snowdon has sight-read many pieces of music. But, Snowdon suggests, these general abilities don't justify attributing particular instances of knowing how which haven't yet been performed.

I think that Snowdon is right that we hesitate to say that the man knows how to get out of a room if he doesn't (yet) know the location of the exit. But this doesn't show that having an ability is insufficient for having know how. It just shows that we need to be more precise when specifying the ability: it's not the fact that the man has a general ability to get out of rooms that matters, but the fact that he has the ability to get out of *this* room. This ability is complex, in that it is built up from smaller parts. To get out of a room, one must be able to move to the door, grasp and turn the knob, push or pull the door, etc. And the man has all of these simpler abilities; he is not paralyzed or disoriented. But until he sees the door, he doesn't have the complex ability of which these are parts. For all he knows before seeing the door, the only escape is through a window, which would require a different set of simpler skills: unlatching the window, pushing it up, lifting himself onto the sill, etc. He can't be

attributed the complex ability, until he knows how to get out of this particular room. It's the *combination* of the simpler abilities that counts as knowing how to get out of the room. Similar remarks apply to the case of sight-reading.

A full defense of knowing how would need to argue that bringing together these simpler abilities to complete a complex task does not involve any propositional knowledge. Sometimes, of course, it will, especially if the task is unfamiliar. If we imagine that the man must climb out a window to escape the room, it's easy to imagine that he thinks through to himself at least some of the steps needed to get out of the window. So, he might silently think: "the window is close enough to the floor that I can lift one leg through at a time rather than pulling myself up by the arms." In such a situation, the knowledge the man employs in forming a plan to get out of the room is clearly propositional. But imagine the original case, in which the exit is a door. In this situation, it seems equally clear that the man could go through the door without any explicit thoughts at all.

To sum up, I think that the proponent of know how can respond to Snowdon's counterexamples. I think that there is a good case to be made in favor of a distinct epistemological category of know how, but a full defense is beyond the scope of this dissertation. In the end, it might be that the debate is at least partly terminological. Both Ryle and the propositionalist agree that there are many situations that don't involve two processes: first the thought, then the action. Sometimes, we just act. The propositionalist denies that the knowing demonstrated by these acts should be distinguished from propositional knowledge. Moreover, the propositionalist agrees with Ryle that some of the knowledge which makes possible certain skills cannot be conveyed linguistically.[20] There are two features which allegedly distinguished know

[20] Ginet says: "It may be that no one can—even that no symbols exist with which it would be possible to—formulate a fully detailed description of the sorts of things one must know how to do in order to ride a bicycle (smoothly) or play a certain piece on the piano (well)" (1975, p. 7).

how from propositional knowledge: know how is acquired by experience and it is ineffable. The know how must be acquired by experience, since it can't be taught. If we broaden our notion of "propositional" to include contents which are ineffable, *pace* Mellor, then we can allow that knowing how is properly conceived as a kind of propositional knowledge.

How would this affect the know how account? On the one hand, because the features that made know how such a good model for explaining Mary's predicament— the experience of seeing red is ineffable and thus knowledge of what having such an experience is like must be acquired through experience— are still present on the propositionalist account, the know how account remains a viable explanation of the central intuition, even if it is really a kind of propositional knowledge. So subsuming know how into the category of propositional knowledge doesn't affect the viability of the know how account as a response to the knowledge argument.

3.2.2 Claims (KH2) and (KH3): The Ability Thesis

I now turn to the claims that together compose the ability thesis: (KH2) Mary gains the abilities to recognize, imagine, and remember colors when she leaves her room and (KH3) the abilities to recognize, imagine, and remember constitute know how. Before defending the ability thesis, it's worth emphasizing again the target of the know-how account. Proponents of this account are not making the metaphysical claim that these three abilities somehow constitute what the experience is like.[21] In other words, the account is not meant to explain what is sometimes called the hard problem of consciousness: the problem of saying why this state feels this way rather

[21] Mellor (1992), p. 13: "Of course none of this helps the know how theory to explain experience in general, i.e. to say what the phenomenal aspect of conscious states of mind is."

than that way or no way at all.[22] The know how account is less ambitious, and claims merely that *knowing* what an experience is like consists in these three abilities.

According to the know how theorist, Mary gains three abilities when she sees red: the ability to recognize, imagine, and remember red experiences. Mellor and Nemirow think that imagination is the primary ability. Mellor thinks it possible for someone to know what an experience is like without having had the experience, and so denies that remembering is necessary. I'll discuss one reason for thinking that imagination does play a special role, but in general I'll follow Lewis in understanding the know how to consist in recognizing, imagining, and remembering. These three abilities are interrelated. Recognition generally requires remembering, since it requires re-identification. Imagination of the sort of interest here—where one is trying to imagine a particular type of experience—requires recognition, since you need to recognize whether or not you correctly imagine it.

Claim (KH2) is generally accepted by those who accept the central intuition, even those who aren't proponents of the know how account. Before leaving her room, Mary has none of these abilities. She certainly can't remember any red experiences, since by hypothesis she has never had one. If she could imagine it while in her room, then presumably she would, in which case she would know what it's like and thus there is no knowledge she is missing. Finally, many have a strong intuition that Mary won't be able to recognize the color red when she leaves her room. Suppose the first colored object Mary sees upon leaving her room is a sample from the paint store. With no contextual clues, i.e. we aren't showing her a banana, it seems very likely that Mary would fail to recognize the color of the sample.

[22] See Chalmers (1996) for a discussion of the hard problem of consciousness. As I'll argue in Chapter Four, it seems that a satisfying account of Mary's situation will have to address the hard problem.

Despite its widespread acceptance, claim (KH2) has been challenged by Janet Levin. She argues that we should distinguish between having a recognitional ability and being able to apply it.[23] Based on Mary's extensive physical knowledge of color, Levin argues that her failure to immediately identify the color sample is indicative not of a lack of recognitional capacity, but of her ability to apply the appropriate color concept. In other words, when Mary sees red for the first time she doesn't gain the ability to recognize red, because she already had this ability while in her room. What Mary learns when she sees is how to apply this recognitional concept.

Levin's story about recognitional capacities seems plausible to me. Mary knows all the physical truths about colors, which will include knowledge of the structure of color space, including the similarity relations that hold among different colors. She'll know that certain colors are considered "warm" and others "cool." Given all that she knows about colors, it seems to me to be an open question whether she has the recognitional capacities that determine how colors are individuated. If she does, then Levin is correct that what Mary learns when she leaves her room is merely the ability to apply her recognitional concept.

Assuming that Levin is right, however, shows only that mere recognitional capacities are not enough for knowing what color experiences are like; *how* we recognize colors is also important for knowing what colors look like. We can imagine a robot designed to make the same perceptual discriminations among colors that human make. For whatever color the robot is shown it can reliably respond with the appropriate color term. We can also suppose that the robot is designed to store its past "experiences" in memory, so that when asked it can recall the color of items it was exposed to in the past. These demonstrations are not sufficient for ascribing

[23] Levin (1986), p. 13. This issue was raised in Chapter One in the discussion of the phenomenal concept strategy.

knowledge of what color experiences are like; knowing what color experiences are like involves recognizing the different colors on a phenomenological basis.

Levin's challenge highlights the importance of imagination for the know how account. I've conceded that it's possible that Mary has the recognitional capacities associated with colors while still in her room, and that when she leaves one of the things she learns is how to apply them. I think that there is a general distinction between having recognitional capacities and exercising them. I might read descriptions of different wines and in this way gain the capacity to recognize what an experience of cabernet franc is like before I've tasted one. Perhaps I never get the chance to taste a cabernet franc. Then it seems possible that I have a recognitional concept, which I never have a chance to apply. It is not plausible, however, to claim that Mary has the capacity to imagine what seeing red is like in her room, but cannot apply it until she has seen red. The distinction between having an ability and being able to exercise it does not fit the case of imagination. If I have the capacity to imagine what a cabernet franc will taste like, I don't have to wait for an opportunity to exercise it. Having the capacity to imagine a situation entails my ability to exercise it.[24]

So Mary doesn't have the ability to imagine what seeing red is like until she leaves her room, and claim (KH2) is partially vindicated. I now want to move to claim (KH3) that the abilities to recognize, imagine, and remember constitute know how. I've claimed that one of the appealing aspects of the know how account is that it introduces a compelling model for the knowledge Mary gains when she leaves her room. We take a category of knowing which poses no challenge for the physicalist—activities like cooking, fishing, riding a bike, eating with chopsticks—and assimilate

[24] Assuming that I am (i) right about know how being a distinct category of knowledge and (ii) that imagining is a kind of know how.

the problematic knowledge. Mary in her room has never had the experience of seeing colors; her position is analogous to someone who has read books on bicycle riding but has never ridden a bike. We wouldn't expect someone who had only read books about bike riding to know how to ride a bike, neither should we expect Mary, who has only read books about colors, to know how to recognize, imagine, and remember the experiences associated with seeing red.

The issue is whether recognizing, imagining, and remembering an experience are genuine cases of knowing how. One obvious objection is that the paradigmatic cases of know how are motor skills. Riding bikes, fishing, tying knots, dancing, and other paradigmatic examples are all abilities which rely to a large degree on manipulation of the body and external objects, e.g. bikes, fishing poles, rope. Recognizing, imagining, and remembering an experience requires no bodily movement, and no manipulation of external objects.[25] The kinds of abilities appear fundamentally different, but I don't think they are. The reason most of the examples involve the body and objects in the world is because most of what humans do (even philosophers) involves moving ourselves and other objects around. But there are examples of know how which don't involve much movement, like knowing how to count, multiply, understand French, or read a map.

A second objection is that Mary can gain the abilities to recognize, imagine, and recall what color experiences are like without any practice. One of the distinguishing features of know how is its mode of acquisition. Know how is not acquired through testimony, because it's very difficult, if not impossible, to linguistically express. The abilities to recognize, imagine, and remember an experience share this feature. If Mary could acquire the skills through testimony, then

[25] This is overstating the case. People often do move when they are trying to remember something; for example, they might stare off into space, grimace, squint, or display some other facial expression, tap their fingers or shake their knee.

she would not need to leave her room. But it's not clear that the mode of acquisition in Mary's case is the same as the mode of acquisition in typical cases of know how. One point of divergence is how long it would take to attain the know how. The standard examples of know how require practice. But recognizing, imagining, and remembering red experiences do not seem to be abilities that Mary would have to practice to attain, nor does it seem that practice will make much of an improvement. Rather it seems that once she has the experience, the abilities immediately follow. This, as most of us can recall, is not what happens when someone first gets on a bike.

There are two ways to respond to this alleged point of disanalogy. One can deny that Mary *knows* what it is like to experience red until she learns to reliably distinguish red things as red and not orange. In other words, one can argue that knowledge of what a red experience is like requires that one know the experience *as* a red experience; recognizing what an experience is like requires categorizing it as a certain type of experience. Like eating with chopsticks or riding a bike, this recognitional ability is a skill that takes time to develop, and it is also an ability that can be improved. If we take the gradual acquisition of abilities as a description of what would happen when Mary is released it might sound unlikely, but this is because she already knows all the functional roles of color experiences. Before leaving her room she knows, for example, that the experience of red is more similar to the experience of orange than it is to the experience of green and that ripe tomatoes typically cause red experiences. Thus it is to be expected that her learning curve will be short. If, the suggestion goes, possession of the recognitional ability requires that one be able to identify properties of one's experience as falling under a certain concept (for example *red*), then, like riding a bike, gaining the abilities will take time and practice.

Tellingly, proponents of the know how account do not argue in this way. This is because they take themselves to be explaining knowledge of what an experience is like when it is not known under any description. They argue that Mary might know what a red experience is like, even if she does not know it *as* a red experience. As David Lewis says,

> One might even know what some experience is like, but not under any description whatever…That is what would happen if you slipped a dab of Vegemite into my food without telling me what it was: afterward, I would know what it is like to taste Vegemite, but not under that description, and not any other non-trivial description. (1988, 99)

As this passage makes clear, Lewis has in mind a situation in which the abilities associated with knowing what tasting Vegemite is like are acquired after a single experience. A parallel case would be one in which Mary comes out of her room and is shown a sample color, and has no contextual clues. Arguably, as long as she was paying attention, upon seeing the color sample Mary would immediately gain the abilities.

It's this sort of case that poses the biggest problem for the physicalist, and so it is important for the know how account to be able to have something to say about such cases. And I think that the know how theorists do have a response. Though most of the cases of know how we've discussed consist of fairly complex abilities, like cooking or riding a bike, there are cases of know how which can be acquired with little practice. If, for example, a person doesn't know how to julienne vegetables she can learn from watching a single demonstration. One can learn the difference between the coo of a mourning dove and the coo of a rock dove from a single comparison of the two calls.

Perhaps not everyone would be able to acquire knowledge in these two cases given only a single demonstration. But most people with experience cooking and bird

watching, respectively, will. Because the experienced cook and birder can use their background knowledge, which will allow them to focus their attention appropriately, on the part of the task that is novel. Like the experienced cook or birder, we imagine what it's like to see a novel color, and similarly assume that we would immediately gain the three abilities to recall, imagine, and recognize it. But Mary is not like us in that she does not have the appropriate background knowledge, and thus it's likely that she would need some practice.[26]

I've argued that the two claims that make up the ability thesis, that Mary gains the abilities to recognize, imagine, and remember color experiences when she leaves her room and that these abilities are kinds of knowing how, can be defended. I think that any resistance to claim (KH3) -- the abilities to recognize, imagine, and remember constitute know how-- is not that these abilities aren't examples of know how, but that this know how doesn't exhaust knowing what it's like. This is the subject of the next section.

3.2.3 Claim (KH4): Recognizing, Imagining, and Remembering is Knowing What It's Like

Many people have pointed out that the ability thesis is subject to counterexamples, which show that having the abilities to recognize, imagine, and remember an experience are not necessary for knowing what an experience is like.[27] Instead of considering each of the abilities individually for ascriptions of knowledge of what an experience is like, let's consider a case where the subject lacks all of the abilities. For example, take someone with an impoverished visual imagination who has lost the ability to form new memories. Call him "Sam." Suppose that Sam is currently being exposed to some color that he has never before encountered in either

[26] Recall the discussion of Molyneux's question in Chapter Two. The empirical evidence indicates that practice is required for the newly sighted to gain the abilities which constitute know how.
[27] See Conee (1985), Levin (1986), Loar (1990), Lycan (1996), Tye (2002).

perception or imagination. Surely he can know what his experience is like while it is occurring despite the fact that in ten minutes he will be unable to recognize, imagine, or recall it.[28]

Because of the limitations of human memory, most of us are in a position similar to Sam's with respect to particular shades of colors. It is frequently noted that our abilities to discriminate shades of colors far outstrips out abilities to remember, recognize or imagine these shades again.[29] Our perceptual experiences are much richer and more determinate than the general concepts we use to describe them. Despite these limitations on memory, we are generally able to know what our occurrent experiences are like. Sam's situation is a genuine possibility.

There are two general strategies for handling counterexamples, one can either accommodate the counterexample or rule it out. We could accommodate the counterexample by broadening our notion of what counts as a recognitional ability. Perhaps all that is required is that Sam is able to recognize and recall the experience as the same experience for as long as the experience is present. This weaker condition is particularly plausible in light of the discussion above. If we assume that Sam has not been told anything about the color he is seeing, e.g. he has not been told its name or some identifying description, then the knowledge he gains is just the sort of knowledge we need to explain.

Unfortunately, I do not think this attempt at accommodation works. Our ordinary understanding of the abilities to recognize and remember crucially involves the reidentification of the items being recognized and recalled. To recognize (or remember) something requires one to think of it as the same thing one has encountered

[28] Requiring memory as part of knowledge of what an experience is like has another odd consequence: when Mary first encounters red (or we have a novel experience) she might know what it is like at the time, but she will have the knowledge only once some time has passed.
[29] See Hardin (1988) and Raffman (1995).

before. By hypothesis, Sam will not be able to reidentify what his experience is like in say ten minutes. Interpreting the ability claim in a way that includes Sam's limited abilities would seriously distort the ways we normally understand the abilities of recognition and recall, to the point that one might then question whether the know how account offers an intelligible analysis of our knowledge of what experiences are like. The know-how account is initially appealing precisely because it analyzes our knowledge of what experience is like in terms of abilities we understand, our folk conceptions of recognition, imagination, and memory. To call the mental processes Sam is engaged in "recognizing" or "remembering" trivializes the proposal; in effect we would be saying, "those abilities one exercises when thinking about red things, whatever they are, are the ones that constitute knowledge of what it's like." The account would lose its explanatory power.

For this reason, I don't think the counterexample can be accommodated by weakening the requirements for being ascribed the abilities to recognize, imagine, and remember. The second general strategy for dealing with counterexamples is to show that they do not need to be accommodated. A defender of the ability claim could argue that without the three abilities, a person could not be said to *know* what the experience was like.[30] Some sort of cognitive process is occurring when Sam introspects his experience, but it is not something that rises to the level of knowledge. Knowledge is stable, and persists through time. This line of argument has the virtue of retaining what was initially appealing about the theory: it relies on ordinary conceptions of the abilities.

And the objection is well motivated. In general, possessing a concept requires that its possessor have the ability to identify or reidentify objects that fall under the

[30] This is similar, though not identical to a suggestion discussed above. There I argued that one response to the objection that Mary would not need practice before acquiring the know how is that she won't have the know how until she can exercise her newfound abilities reliably.

concept.[31] If this is accepted as a condition on concept possession, it means that until a person consistently applies the concept correctly, she cannot be ascribed the concept; hence, she doesn't have the knowledge. This consistency seems to be a hallmark of our ordinary knowledge ascriptions; until some level of consistent behavior is reached the situation is described as one of learning, not knowledge. Think, again, of the abilities to ride a bicycle. If four out of five times you fall off the bike, it would be wrong to claim you know how to ride a bike.

Though well motivated, the objection is misguided. For whether or not we call what Sam gains 'knowledge' isn't the crucial issue. We're trying to explain what happens to Mary when she leaves her room. For this purpose, what really matters is that Sam is able to entertain the thought that this is what my experience is like, while lacking the abilities to recognize, imagine, and recall the experience. If this is a coherent possibility for Sam, then it will also be a possibility for Mary. Mary once released will be able to entertain a thought she could not have entertained before and that cannot be explained by appeal to the three abilities.

The counterexample shows that the abilities aren't necessary for knowing what an experience is like. I think it also shows, as I discussed in Chapter Two, that the central intuition isn't an intuition about knowledge, but about the kinds of thoughts Mary cannot have. For the purposes of explaining the central intuition, it doesn't matter that it is *knowledge* Mary gains when she leaves her room.

The above discussion indicates that there are two types knowledge that Mary can gain, and the ability thesis can explain one and not the other. The ability thesis is compelling if we are trying to explain our general knowledge of what experiences are like. I know what the experience of smelling lilacs is like because I can recognize,

[31] Sean Kelly (2001a), for example, argues for a re-identification condition for demonstrative concepts. Diana Raffman (1995) also questions the coherency of an ability to identify a thing but not re-identify it.

imagine, and remember what lilacs smell like. Having these three abilities is sufficient for ascribing knowledge of what the experience is like to me. Furthermore, it seems that this knowledge is a kind of knowing how, since it seems to crucially depend on an ability to imagine or recall the smell of lilacs even when the lilacs aren't present.

But this know how is different from the kind of knowledge I have when I'm actually smelling lilacs and attending to what it's like. In that case, as Sam's unfortunate situation demonstrates, none of the abilities are necessary. The knowledge I have when I'm attending an occurrent experience of smelling lilacs is very different from imagining or remembering what it's like to smell them. When I attend to an occurrent experience, the knowledge I have is much richer. Not only are there are more features of the occurrent experience, but these features have a richer presence. My imagining of the experience is but a faint copy of the actual experience. It's the knowledge of occurrent experiences which is missing from the know how account.

3.3 An Alternate View

3.3.1 Stanley and Williamson's Proposal

I finish this chapter with a brief discussion of a way to incorporate the insights of the know how account, while abandoning the claim that Mary's new knowledge isn't propositional. Jason Stanley and Timothy Williamson (2001) argue against distinguishing between know how and propositional knowledge, and argue in support of the claim that know how should be assimilated to propositional knowledge. I'm not concerned here with evaluating their arguments against the cogency of the distinction, since even if there is a separate category of know how it seems that some of Mary's new knowledge is propositional. Rather I want to examine Stanley and Williamson's

positive proposal concerning the knowledge that Mary gains when she leaves her room and learns how to imagine red. [32]

Stanley and Williamson claim that Mary's ability to imagine red consists in her knowledge of a (Russellian) proposition involving a way of imagining red which is thought about under a practical mode of presentation. "Ways" here are properties of token events. Here is their proposal:

> *x*'s knowing how to imagine red amounts to knowing a proposition of the form '*w* is a way for *x* to imagine red,' entertained under a guise involving a practical mode of presentation of a way. (2001, 442)[33]

For humans with normal color vision, our knowledge of how to imagine red is explained by our knowledge of a proposition like 'this is a way for me to imagine red' when we think about the way under a practical mode of presentation. This treatment can be extended to the abilities to remember and recognize.

It might be helpful to think about Stanley and Williamson's treatment of a different example. Consider the sentence: "Sally knows how to ride a bike." According to Stanley and Williamson, this sentence is true if Sally can entertain a proposition of the form '*w* is a way for Sally to ride a bike' under a practical mode of presentation. The requirement that the proposition be entertained under a practical mode of presentation is necessary to distinguish it from the following kind of case. Assume Sally doesn't know how to ride a bike, but she has been watching bike riding on television. Pointing to the image on the screen she says 'that is a way for Sally to ride a bike' and assume that what she sees on the screen would be a way for her to ride a bike. In this case Sally is entertaining the way under a demonstrative mode of presentation, rather than a practical one, and this can be true in cases where Sally doesn't know how to ride a bike.

---

[32] See Koethe (2002), Rumfitt (2003), and Noë (2006) for critical discussion of Stanley and Williamson's arguments.
[33] Cf. Ginet (1975), p. 6.

Clearly, it is the notion of "a practical mode of presentation of a way" that is doing most of the work. Analogously to the case above, Mary can, while still in her black and white room, entertain a proposition of the form '*w* is a way to me to imagine red.' For example, she can entertain the proposition that the way ordinary humans imagine red is a way for me to imagine red. Or, given the right kind of brain-viewing apparatus, she could demonstrate the brain event occurring when someone outside the room imagines red. What she cannot do is think about ways of imagining red that involve practical modes of presentation.

3.3.2 The Proposal Fails the Explanatory Constraint

I'm not sure if Stanley and Williamson intend the application of their positive proposal as a response to the knowledge argument, but for the purposes of this discussion I am going to treat it as a genuine response.[34] For the strategy of appealing to a new mode of presentation to explain what Mary learns is a popular one and I want to show its *explanatory* failure when applied to the knowledge argument. It might be, for all I'll have to say, that Stanley and Williamson's analysis of knowing how is correct. The application of their analysis to the knowledge argument, however, fails as an explanation of the central intuition.

On the Stanley-Williamson proposal, Mary lacks knowledge of what red things are like because she is unable to think about a way of imagining red under a practical mode of presentation. Thinking about a way under a practical mode of presentation is a technical notion. Unlike the abilities to recognize, imagine, and remember, we don't have any folk conception of what it is to think about a way of imagining red under a practical mode of presentation. And, as they acknowledge, giving a nontrivial

[34] That is, Stanley and Williamson might take their theory to explain what Mary learns, but not think that this is all that needs to be said on behalf of the physicalist.

characterization of a practical mode of presentation is a "substantial philosophical task" that they do not undertake (2001, 429). But without such a characterization, there is no explanation for why Mary cannot acquire such a way while still in her room.

More troubling, their application of their view to the knowledge argument fails to meet the explanatory constraint. After taking themselves to have shown that all know-how is propositional knowledge, Stanley and Williamson mention a possible fallback position for the know how account, a version of Levin's proposal. Perhaps Mary does know how to imagine red while still in her room, but until she sees red for the first time she lacks the ability to apply this knowledge.[35] But, Stanley and Williamson claim, this is "absurd" (2001, 443). Why, they ask, if she knows how to imagine an experience of red while still in her room is she unable to imagine the experience? Knowing how to imagine red entails having the ability to imagine red, Stanley and Williamson claim.

The commonsense reason for why Mary cannot imagine such an experience is that she has never instantiated the experience herself. Until Mary has had an experience of red, the intuition goes, she will be unable to imagine (or recognize or recall) the experience. According to Stanley and Williamson, Mary must be able to instantiate the way herself before she can know what seeing red is like. This is necessary to distinguish the practical mode of presentation of a way from the demonstrative mode of presentation of a way. But then the practical mode of presentation amounts to knowing a way to instantiate a way to imagine red. And this is merely to claim that one must have the experience of seeing red in order to know how to imagine red.

[35] This position is defended in Levin (1986).

In other words, if we take Stanley and Williamson to be offering a response to the knowledge argument, then, I have argued, to be adequate the response needs to meet the explanatory constraint; that is, it should explain why Mary cannot have knowledge while still in her room. But Stanley and Williamson are appealing to the central intuition—that Mary lacks knowledge while still in her room—to explain why she can't have the practical mode of presentation of a way of imagining red.

Conclusion

I argued that an adequate response to the knowledge argument must meet the explanatory constraint: it must provide some noncircular explanation for why Mary must leave her room to learn what red looks like. I argued that the know-how strategy meets this explanatory constraint by offering a model for Mary's new knowledge. The know-how strategy points to a class of knowledge which seems less mysterious and tries to assimilate Mary's new knowledge to this more ordinary kind. It does this by pointing to two similarities between ordinary examples of knowledge and the knowledge Mary gains when she leaves her room: the ineffable nature and the mode of acquisition.

Unfortunately, while meeting the explanatory constraint this proposal also met with a counterexample; it seems that Mary can have knowledge of what an experience is like without having any of the three abilities. We then looked at an alternative proposal, one that has been quite popular. For all I've said, Stanley and Williamson's proposal—that Mary gains a new mode of presentation—might turn out to be true, but it assumes rather than explains the central intuition. In the final chapter, I return to tackling the central intuition head-on, in particular, making sense of the notion that perceptual experience is ineffable.

Chapter Four
The Ineffability of Perceptual Experience

Introduction

I argued in the last chapter that what Mary fails to grasp while still in her room is not adequately explained by the ability hypothesis. Before offering my own explanation for why Mary can't know everything about colors without seeing them, we need to explore in more detail what, precisely, it is that Mary learns when she sees (say) yellow for the first time.

Let's begin by discussing what Mary *can* know before leaving her room.[1] Before leaving her room Mary can believe, for example, that ripe tomatoes are red, that lemons are yellow, and that Juan's favorite color is chartreuse. However, the beliefs that Mary has before leaving her room are different in kind from the beliefs that she'll acquire after seeing colors herself. To see this, imagine the following situation. Mary, while stuck in her room, is told by Juan that his favorites color is chartreuse. On the basis of his testimony, she comes to believe that Juan's favorite color is chartreuse. She and Juan share the belief that Juan's favorite color is chartreuse. Let's call this belief that Juan and Mary share a *nonphenomenal belief.*

In addition to the belief that Juan shares with Mary, Juan has a belief that Mary cannot have before leaving her room, a *phenomenal belief* which involves the way chartreuse appears to him. The difference between the two beliefs is illustrated by what might happen when Mary sees colors for the first time. Suppose we show her an array of colors and ask her to pick out Juan's favorite color. She chooses a burnt orange color. Because Mary has heard so much from Juan about the aesthetic virtues of chartreuse, she is confident in her choice and forms the phenomenal belief that

[1] The following discussion is indebted to Nida-Rümelin (1995).

Juan's favorite color is burnt orange. In other words, Mary mistakenly believes that the concept CHARTREUSE applies to burnt orange. Here we can see that although Mary correctly believes that Juan's favorite color is chartreuse, there is a belief that she is missing while still in her room: a belief about what chartreuse looks like.[2]

One could argue that her acquiring this mistaken belief about what Juan's favorite color looks like shows that we shouldn't have ascribed to her the nonphenomenal belief that Juan's favorite color is chartreuse. In other words, it could be argued that Mary cannot have any beliefs which would be expressed using color words until she has been out of her room. I am inclined to be more ecumenical about belief possession. It seems to me perfectly acceptable to attribute to a person blind from birth some beliefs about colors. Someone who has never had a visual experience can have the belief, say, that his car is blue based on the testimony of others.[3]

Before leaving her room, Mary can't have the phenomenal beliefs, though she can and does, have beliefs *about* the phenomenology. For example, while in her room she might have been told that blues and greens are cool colors whereas yellows and reds are warmer colors. These are phenomenological facts about the way the experiences of seeing blue shades and red shades feel to the typical human observer,

[2] Our nonphenomenal beliefs about colors are dependent on the phenomenal beliefs we have about colors. This dependence, I take it, is obvious. If humans were configured differently and as a result were completely colorblind and incapable of enjoying the phenomenology associated with color experiences, then the typical person wouldn't have any beliefs about colors at all. Our color categories and color vocabulary are determined by the qualitative features of colors; we distinguish one shade from the next by the way the shade looks to us, and we make judgments about the similarity relations which hold between shades by how similar (or different) the compared shades look.

[3] It's a difficult question to know just what's necessary for ascribing a belief. Should we ascribe to a three year old the belief that Chardonnay is drier than Riesling because her parents told her so? What about an adult teetotaler who knew nothing about wine? I'm inclined to count these as cases of genuine belief, but others might not be. (Part of this depends on how much one wants to defer to the community for assessing speaker meaning, and the tricky question of how to individuate the community—as the adult teetotaler illustrates.) Since the cases I'm interested in involve subjects with many beliefs about colors, I'm not going to worry about these kinds of cases.

and it seems plausible to think that Mary can know these facts in the same way she knows that lemons are yellow, that is, by testimony.[4]

There's a way to interpret Mary's situation so that it is very similar to the Molyneux case. Remember in Chapter Two I distinguished between the question of whether the blind man will be able to visually distinguish the two objects from each other, and the question of whether the blind man would be able to correctly identify the one object as a cube and the other object as a sphere. We can also distinguish between two parallel questions here. First, will Mary have the right cognitive and physiological equipment to see a difference between red and blue colors? (She has, after all, grown into adulthood never having been exposed to color.) Second, assuming that Mary will be able to distinguish red and blue, will she be able to correctly identify red things as red and blue things as blue?

Though dubious on empirical grounds, I'll assume that the answer to the first question is yes. But I am also willing to allow for the possibility that the answer to the second question is yes. That is, as Leibniz argued about the blind man, it seems possible that Mary could know enough about colors to accurately identify a ripe tomato as red on leaving her room for the first time. Perhaps Mary could extrapolate from what she knows about colors to correctly identify her experiences. In place of the tactile knowledge of the blind man, Mary has scientific and descriptive knowledge. It seems possible that Mary, as some philosophers have argued about the newly sighted Molyneux subject, could infer from her descriptive knowledge that what she is now seeing is (say) a red tomato. However, even if Mary could correctly apply her color concepts, I don't think this is a counterexample to the claim that Mary must have experiences of colors in order to know everything about them. What this example

[4] More could be said about the distinction between phenomenal and nonphenomenal belief, but the basic idea is intuitive. See Nida-Rümelin (1995) for a detailed defense of the distinction between phenomenal and nonphenomenal believing. See also Chalmers (2003).

shows is that the relevant issue is not whether Mary can correctly identify colors when she is released from her room.[5] Even if she does know how to correctly apply the concept RED, she will still be surprised by what colors look like. There will still be something she learns, which she couldn't know while in her room.

Any reluctance to concede that there is something Mary doesn't know before leaving her room is driven by Mary's unusual intellectual situation. As Jackson imagines Mary she is omniscient with regard to color science; she knows all the physical and functional truths there are to be known about colors. As normal human subjects, it's hard to imagine what kinds of mental powers we would have if we were in Mary's situation. So to simplify intuitions, I'll assume in what follows that Mary is just an ordinary person, with ordinary mental powers, in an extraordinary situation. Since, as I explained in Chapter One, my primary consideration is not defending physicalism but exploring the central intuition, it doesn't make a difference whether Mary knows all the physical truths.

There are two reasons for simplifying intuitions in this way. First, it allows us to focus on the central intuition of the knowledge argument: the intuition that seeing colors is necessary for acquiring certain types of knowledge. Second, it doesn't rule out the possibility of a materialist response to the original knowledge argument. As I just mentioned, the reason it is so difficult for us to imagine Jackson's Mary is because she is so unlike us. Acceptance of the central intuition is based on what we know about actual cases of human perceptual experiences (especially, it could be argued, our own). Since the central intuition is based on our understanding of actual cases, it seems reasonable to focus on a situation that is causally possible. And if the proposal I offer for why ordinary Mary learns something new when she sees colors is at least

[5] Recall the discussion in Chapter One of the missing shade of blue. I also argued there that the primary issue is not whether the recognitional abilities are present.

compatible with physicalism, then it will count as one of the (many) possible responses to the knowledge argument.

So what does Mary learn when first shown, for example, a lemon? It seems she learns two distinct things. She learns something about the world - that the lemon has a property of appearing a certain way. And she learns something about her experience of the world - that her experience of yellow things has a certain property.[6] In other words, she acquires two beliefs: a belief that the lemon appears in a particular way, namely, yellowy, and a belief that there is something the experience of seeing yellow feels like and it's like the experience she is undergoing. The first belief attributes a property to the lemon, and the second attributes a property to the experience of seeing the lemon. (Even if Mary has grounds for thinking that the lemon is not yellow, e.g. she doubts the reliability of her vision, she will still have a belief about the way the lemon appears to her and what that experience is like.)

How does Mary acquire these beliefs? Imagine a newly released Mary looking at a lemon.[7] By hypothesis, there are three occurrent states Mary undergoes when she sees the lemon: the visual experience which presents the lemon as yellow, the belief about the way the lemon appears, and her belief about what the experience of seeing a lemon is like. And the phenomenology—the way the lemon appears to the subject—is the same for each state. Mary's visual experience presents the lemon in a particular way, i.e. yellowy, and her belief about the way the lemon is presented (that it looks yellow rather than red, and this shade of yellow rather than another) and her belief about what seeing a lemon is like are based on the content of this experience.

---

[6] I mentioned in Chapter Two that Locke clearly has in mind the latter kind of belief. See Byrne (2002) for a contemporary discussion of this distinction. On certain accounts of the nature of perceptual experience, this second belief will be false. But pre-theoretically, it's natural to attribute properties to experiences, e.g. experiences can be painful, exhilarating, awkward, tedious, etc.

[7] I will assume throughout the paper that the cases of interest are those in which the subject is paying attention to either the object of her experience or the experience itself. I will not be concerned with aspects of perceptual experiences to which the subject is not paying attention.

The three states are different. A person can have a visual experience of a lemon without any corresponding beliefs, if, for example, she isn't paying attention to her surroundings. And we typically have beliefs about lemons without having any corresponding beliefs about what our visual experiences are like; that is, in our ordinary life we don't bother to reflect on each experience and think about its phenomenology. Conceptually, having the experience doesn't entail having the first-order belief, and having the first-order belief doesn't entail having the second-order belief. (The converse is not true. Having the second-order belief depends on having the first-order belief, which in turn depends on having the experience.) Despite these differences, these three occurrent states seem to share certain presentational features—how the lemon appears is the same in each of them—and I will call this presentational aspect the *phenomenological aspect.*

As I pointed out above, not all beliefs about possible objects of perception have a phenomenological aspect. Mary's belief that lemons are a good source of antioxidants, for example, doesn't have any phenomenological content; this is an example of a non-phenomenal beliefs discussed above. One of the goals of this paper is to explain the difference between this kind of belief and the kind of belief Mary has on her release. At first glance, the answer seems obvious: Mary's beliefs about how lemons look are based on her perceptual experiences, and her beliefs about lemons' nutritional value are not. But this answer isn't very illuminating. What I want to know is *why* knowledge of certain truths is available only via experience. Why can't Mary understand truths with phenomenological content in the same way she understands other truths? Is it because of the nature of experience, limitations on the subjects of experience, the structure of natural languages, or some combination?

In this chapter, I argue that Mary cannot understand these truths because the phenomenological aspect of the content of the belief is ineffable, by which I mean that

this aspect is not linguistically expressible. If the content of perceptual experience were nonconceptual, as many philosophers have argued, then this would explain why it is ineffable. I explore and reject this suggestion in the second half of the paper. But first it will be helpful to take a closer look at the notion of ineffability. In the first section of this chapter, I outline the basic structure of the proposal and introduce two distinctions, the first between types of ineffability, and the second about types of theories. In section 2, I examine two recent proposals about how we should understand the ineffability of perceptual experience.[8] In section 3, I explore the relation between nonconceptual content and ineffability, and in the fourth section I explain and defend my own phenomenological proposal as an explanation of the ineffability of perceptual experience.

4.1.1. The Ineffability Proposal

Let's look again at the central intuition, the intuition that Mary must leave her room to learn certain truths about colors. This intuition is an instance of a general principle, which I will call the *Experience Thesis*:

> (Experience Thesis) To have a belief with phenomenological content P, one must have (or have had) an experience with phenomenological content P.

The Experience Thesis lies behind Jackson's argument; without this intuition his thought experiment doesn't work. People have the intuition that Mary learns truths about color perception after being released from the room *because* they believe the Experience Thesis is true. There are certain beliefs that are unavailable to people who have not had the requisite experiences. Mary cannot know what it's like to see yellow or what it is for a bead to appear yellow, until she has the experience of seeing

[8] The suggestion that perceptual experience is partly ineffable is also raised in Byrne (2002), Hellie (2004), and Thau (2002).

something yellow.[9] But, as the Experience Thesis makes clear, Mary can't have the knowledge, because she can't form the relevant belief; that is to say, Mary's difficulty with acquiring knowledge is not like the standard problem of not having adequate justification, but with having the relevant mental state.

Like many others, I find the Experience Thesis compelling. I propose the following explanation for why it is true in Mary's case:

> (Ineffability Proposal) Mary can't have certain beliefs about colors before she has experienced them, because the phenomenological content of visual experiences is ineffable.

Let me emphasize, again, that Mary could be told *about* the phenomenological content. She could be told, for example, that the yellow of the bead is like the yellow of the sky. It's only those beliefs *with* phenomenological content that she can't have without the relevant experience. In other words, Mary cannot be told the phenomenological content, because this content cannot be fully expressed linguistically. If Mary cannot be told the content, then she can't form the relevant beliefs (necessary for having knowledge) for a simple reason. We have only a few ways of gaining knowledge about the world: being told, having experiences, or extrapolating from the knowledge gained in these ways, conceptually or imaginatively. Since Mary has never had an experience of seeing anything colored, the latter two can be eliminated as ways she might come to form the relevant belief.[10] So the only way left for Mary to learn about the phenomenological content is to be told. But, if the Ineffability Proposal is true, then she can't be since phenomenological content is not linguistically expressible.

---

[9] Maybe you don't have to actually see a yellow thing; perhaps all you need is to hallucinate a yellow thing. I'm going to ignore this complication.

[10] This is a bit too quick. Daniel Dennett, for example, argues that Mary could extrapolate from her knowledge of the physical truths knowledge of how yellow will appear. This is addressed in more detail in the next section.

The Ineffability Proposal is offered as a proposal about *why* the Experience Thesis is true; that is, the Ineffability proposal meets the explanatory constraint discussed in Chapter Three. Mary cannot form beliefs with the phenomenological content associated with seeing yellow things without having the experience of seeing something yellow, because the information she needs to form the relevant belief cannot be linguistically expressed to her. Nothing that we could verbally communicate to her would enable her to form a belief with the appropriate phenomenological content. Until she has had her own experience of seeing yellow, there are certain truths about color perception that will be beyond Mary's reach.

4.1.2 Weak and Strong Ineffability

In this section, I want to discuss two reasons for why the phenomenological content of an experience might be ineffable. First, the ineffability might be the result of cognitive limitations on the human mind. Second, the ineffability might arise from the phenomenological nature of the experience itself. The former I call *weak ineffability* and the latter I call *strong ineffability*. In what follow I explain this distinction and argue that only strong ineffability will explain why Mary can't learn certain truths about colors before she's had color experiences.

Weak ineffability arises from empirical limitations that affect our ability to describe our experiences. For example, there are well-known limitations on our abilities to accurately recall shades of colors, even though we are capable of discriminating between the shades.[11] Show an ordinary human subject two very similar shades of green, G1 and G2, between which she can discriminate when the shades are presented side by side. When presented with G1 by itself, she is unlikely to be able to tell whether it is G1 or G2 with any degree of accuracy. It follows from this

[11] See Hardin (1988), Raffman (1995).

inability that she will not be in a position to describe with precision visual scenes that are not occurrent. It also follows that she will not be in position to describe an occurrent scene in a way that distinguishes it from a nearly identical scene.

In addition to constraints on memory, there are limitations on how quickly we can process information provided by our visual systems. Imagine how long it would take to describe in complete detail your current visual experience, and then imagine trying to do that for a day's worth of experiences. Since we are never in situations in which the environment is completely unchanging, the task is even harder. Environments change before being entirely processed. Given these limitations on processing speeds and the constraints of memory, we are unable to describe every aspect of a given experience.

So, there are memory, processing and time constraints on our ability to describe our experiences in full.[12] When these constraints are the source of ineffability, it is weakly ineffable. More precisely, the phenomenological content of an experience is *weakly ineffable* if (a) we cannot describe it in words and (b) if constraints of memory, processing and time were removed we could compose a lengthy description that could be understood by someone who was not undergoing, and had not undergone, a phenomenologically identical experience. If the phenomenological content of our experiences is weakly ineffable, then this failure to linguistically express the phenomenological content of our experiences is explained by cognitive limitations of the human mind, and not by anything about the nature of our experiences.[13]

[12] This is not an exhaustive list. There are other types of computational constraints, discussed in section 2 below.

[13] Dennett (1988) seems to have this type of ineffability in mind in his deflationary discussion of qualia. I agree with him that weak ineffability poses no threat to physicalism.

Before applying this to Mary's situation, let me illustrate the distinction between weak and strong ineffability with a different example. Imagine a being with a visual system phenomenologically like ours, but without any computational constraints on memory and processing.[14] Call him "Compy." Imagine that Compy is looking out the window and talking to a similarly unconstrained Friend on the telephone. Compy is enjoying a particularly beautiful sunset, which is phenomenologically identical to the experience a human looking out the same window would have, and decides to describe it to Friend. Friend is hundreds of miles away, so in his description of his visual experience, Compy can't rely on Friend's enjoyment of the same experience (so his use of demonstrative expressions wouldn't help). If it is possible for Compy to fully capture the phenomenological content of his visual experience in a way that Friend will be able to fully understand, then the phenomenological content of experiences is merely weakly ineffable.

By contrast, strong ineffability results from the phenomenological nature of visual experiences, rather than from cognitive limitations of the experiencers. If the content of experience were ineffable in this sense, then even the phenomenological content of a creature with only a single type of experience, like Hume's oyster, would be linguistically inexpressible.[15] Even if the constraints on human memory and human processing were removed, we would still be unable to express fully the contents of our experiences. If it is impossible for Compy to linguistically express all that he sees when he looks out the window, then phenomenological content is not weakly ineffable, but strongly ineffable.

The phenomenological content of an experience is *strongly ineffable* if it is impossible for it to be fully captured by a linguistic description. If the

[14] Thanks to Carl Ginet for suggesting this example.
[15] Hume imagines the following: "Suppose the mind to be reduc'd even below the life of an oyster. Suppose it to have only one perception, as of thirst or hunger" (1739/2001, p. 399).

phenomenological content of our experiences is strongly ineffable, then there is some property of the phenomenological content, rather than some lack of ability on the part of the subject of the experience, which explains why it cannot be linguistically expressed.

The claim that the phenomenological content of perceptual experience is strongly ineffable is the more interesting one, because it is a claim about the nature of visual experience itself, and not a claim that results from contingent facts about the abilities of subjects of visual experiences. It is also of immediate interest to me, because only a version of the Ineffability Proposal involving strong ineffability would successfully explain the Experience Thesis—the claim that we need to have an experience with phenomenological content before we can have a belief with the same phenomenological content.

To see why strong ineffability is necessary for a satisfying explanation of the Experience Thesis, consider Mary's predicament. She doesn't need us to describe the entire phenomenological content of a typical visual experience, but only the phenomenological content associated with a single shade; computational limits would not prevent our doing this.[16] If computational limits were the cause of the ineffability of visual experiences, then Mary should be able to form the relevant beliefs about the color of the bead before leaving the room. The Experience Thesis, and the central intuition it is meant to explain, tells us that this is impossible. So if our experiences were merely weakly ineffable, then their ineffability would not explain the Experience Thesis; that is, the Ineffability Proposal would fail.

[16] This isn't entirely accurate. The computational limits discussed above would prevent us from describing a shade in such a way that Mary could pick it out from a nearly identical shade. But the problem here is worse: we couldn't describe any shade in way that allowed Mary to form a belief with the right sort of content. Thanks to Anne Nester for bringing this point to my attention.

For further illustration, imagine Compy's Friend in Mary's situation. Friend has never seen anything yellow, but is not subject to the computational limits of Mary. Compy tries to describe the phenomenological content of the visual experience Friend will have when she leaves the black and white room. Will Friend be able to form a belief with the phenomenological content Compy describes? My intuition is that she will not, and thus that the Experience Thesis is true of creatures that enjoy experiences with phenomenological content like our own, but who aren't subject to the same computational limits.

As I mentioned in Chapter Three, I am sympathetic to Dennett and Churchland's claim that we, as humans with ordinary mental powers, aren't good at imagining omniscient Mary's situation. And so I am sympathetic to the claim that we cannot determine, by relying on intuition, whether or not Mary's inability to form certain beliefs is the result of computational limits on the human mind. This is why I attempt, in the following sections, to find some feature of phenomenological content that would explain why phenomenological content is strongly ineffable. As it stands now, what we have is a clash of intuitions. To move past this stalemate will require finding some actual evidence for strong ineffability.

To sum up, in order for the Ineffability Proposal to have a chance at providing an explanation for the Experience Thesis, it must be that phenomenological content is strongly ineffable. Any argument which supports merely weak ineffability will not count as evidence for the Ineffability Proposal. In the next section, I look at two different theories of ineffability, which seem to explain merely the weak variety.

4.2 Two Accounts of Ineffability

Let me make clear again my use of the word "ineffable." As I just mentioned, the ineffability of perceptual experiences merely means that these experiences are not

(entirely) linguistically expressible. In particular, being ineffable does not mean being unknowable, nor does it entail any claims about knowledge. As we saw in Chapter Three, it seems very plausible to think that we can have knowledge that is not linguistically expressible.[17]

In the following two sections, I examine two theories of ineffability. The two theories offer different kinds of explanations, what I call *representational* and *perspectival* respectively, and I discuss them in part as a contrast to my own *phenomenological* explanation. The first theory, proposed by Diana Raffman, is a theory of musical ineffability, which might appear somewhat removed from my own interest in ordinary, visual experience. However, she has offered the most complete account of ineffability I am aware of, and I use her account of musical ineffability to represent a general cognitivist approach that can be applied to the ordinary cases of visual perception. The second theory I discuss is William Lycan's. Lycan's theory of ineffability is a result of his higher-order theory of perception, and has the virtue of being offered as a partial response to the knowledge argument. In the final section of the chapter, I will offer my own suggestion for the ineffability of perceptual experience.

---

[17] It might be worth pausing a moment here over a potential ambiguity over what thing has the property of being ineffable: is it the property of the object being represented or a property of the representation? There are times when people use the word "ineffable" in the former sense, to designate a property of an object. So, for example, sometimes it is claimed that certain of God's properties are ineffable, or that the object of certain mystical experiences is ineffable. And sometimes the word is used in the latter sense, describing the representation of an object. For example, our experiences, beliefs, or concepts of God might be called "ineffable."

I've been using "ineffable" in this second sense, as a property of the representation, in this case a property of visual experience. Perhaps though there is something about the object of these representations that is ineffable. It does seem more difficult to describe the phenomenology of the so-called secondary qualities like tastes, sounds, and smells, than it does to describe the phenomenology of the so-called primary qualities, like shape or movement. So perhaps there is something about the nature of properties (and not merely their representations) like color, which explains why representations of these secondary properties are ineffable. Exploring this issue would take us too far a field, into specific accounts of colors and the viability of the distinction between primary and secondary qualities, so in what follows ineffability should be understood as a property of mental states or their constituents.

4.2.1 Raffman's Representational Account

In her 1993 book, *Language, Music, and Mind*, Diana Raffman offers a theory designed to explain what is "one of the most deeply rooted convictions in modern aesthetics: our knowledge of artworks is, in some essential respect, ineffable" (1993, 2). Raffman's particular concern is with musical ineffability, and the cognitivist theory she eventually offers is built upon empirical claims about how we process music. Given both the object towards which Raffman's theory is directed, musical ineffability, and her method, which is to examine the psychological representations we use when listening to a piece of music, her theory might seem to be an odd place to look for an explanation of the ineffability of visual experience. However, I think that Raffman's account is relevant to my discussion of the ineffability of visual experience for (at least) two reasons. First, even though her theory relies on particular empirical features of our auditory representations, it is reasonable to suppose that something analogous can be said for each of the four other senses. Second, Raffman's account is representative of what I take to be one of the three main approaches to the problem of ineffability; her account illustrates *the representational view*.

I must issue a quick disclaimer before discussing the details of Raffman's account. There is no way to explain her theory without getting into a few of the technical details. A full critical examination of her theory would require expertise in many areas in which I am not an expert: musical theory, psychology of music, and aesthetics, for example. Luckily, this doesn't matter for my purposes. I am not interested in offering criticism of Raffman's particular claims about why musical performances are ineffable. I am interested in seeing how her general theory could be applied to visual experience, and what such a theory would imply about the ineffability of visual experience.

Raffman begins by distinguishing three types of ineffability: structural, feeling, and nuance. *Structural ineffability* is so-called because it is ineffability which is associated with the structure of the piece of music. Raffman distinguishes between two types of structural features, global and local. An example of a global structural feature would be the phrase boundaries in a piece of music. An example of a local feature is rhythm. Raffman associates structural ineffability with the global structure. There are two reasons for this. First, the trained musician can report the local structures, but struggles to explain why she slows down during a certain passage (which is in indication of where she locates the phrase boundary).[18] Second, musicians are consistent in their reports of local structures both inter and intrasubjectively, but not very consistent at the global level.

Structural ineffability is *in principle* linguistically expressible. Sometimes, according to Raffman, what was ineffable to a musician becomes effable when she studies the score (1993, 32). And, according to Raffman, the global structures are no different in kind than the local structures (1993, 31). So while the trained ear sometimes fails to be able to exhaustively report the structure of a piece of music, the structure is in principle effable. The reverse is also true; a musician can acquire knowledge of a musical structure by reading or being told the score, which allows her to imagine what the piece sounds like. Raffman is clear that this type of musical knowledge is not merely descriptive, but involves sensory-perceptual states; in other words, the musician reconstructs an auditory image from reading or hearing the score. The result of this process is that she knows what the music will sound like (1993, 38).

A musician can acquire structural musical knowledge from a description, but only if she has first had experiences of the right sort. The sensory-perceptual concepts

[18] Raffman assumes throughout her book that the subject of the musical experience is someone with training, because she wants to rule out the ineffability being the result of ignorance. I will also assume a musically knowledgeable subject in what follows.

she uses in the reconstruction are recognitional; she must have heard the right sounds in order to form the appropriate representations. In other words, no one deaf from birth can know a piece of music. This feature of our musical knowledge Raffman calls *feeling ineffability* (1993, 39-42). Feeling ineffability is more powerful than structural ineffability. Even if we were able to exhaustively describe the structure of a piece of music, our musical knowledge would still remain feeling ineffable.

Feeling ineffability is related to the sorts of ordinary, everyday ineffability that I am concerned with, but Raffman thinks that it is especially powerful as a feature of music. She thinks that the feeling ineffability of music, as compared to our ordinary perceptual experiences, is especially good at grabbing our attention.[19] She attributes feeling ineffability, at least in part, to the essentially occurrent nature of musical performances. While Raffman believes that feeling ineffability should be recognized by any adequate account of musical ineffability, she also claims that this kind of ineffability is not adequately addressed by the cognitive account she offers (4). However, as we'll see, the distinction between feeling and nuance ineffability is not hard and fast, and some of her explanations of nuance ineffability seem appropriate to apply to feeling ineffability, too.

The third type of ineffability that Raffman distinguishes is *nuance ineffability*. Nuance ineffability is the "heart of [the] theory," and the type of ineffability that Raffman thinks other writers have had in mind when discussing musical ineffability (4, 99). The details of her explanation of nuance ineffability are complicated, but the basic picture is fairly simple. Raffman argues, on the basis of empirical evidence, that there is a level of auditory representation that we are unable to type-identify. That is,

[19] And she offers the following interesting suggestion for why this might be so (1993, 61). Raffman argues at length that music shares with language a kind of grammatical structure (see Chapter 2). Because music shares this feature with language, when we hear a piece of music we expect a semantics; when we don't get it, we are surprised. Thus, the ineffability of music is continually surprising to us, which makes us take notice of its ineffable character.

there are mental representations which we fail to categorize, and thus are unable to verbally report, even though the representations are consciously accessible.

The "nuance" in nuance ineffability refers to the features of our sensory-perceptual knowledge that are nonstructural, for example, pitch. Nonstructural features are features that are not written into the score, and are the kinds of features which allow for different interpretations of the same piece of music. Thus, nuance ineffability is stronger that structural ineffability; reading the score won't help you articulate nuance ineffability. (The relationship between nuance and feeling ineffability is trickier to pin down. This is discussed below.) Raffman claims that our inability to categorize these nuances is central to how we hear the music: "it is plausible to suppose that if N-interval schemas were activated in the perception of a musical performance, every nuance therein—every determinate pitch and interval we can hear—would be *recognized*, and the melody (for one) would vanish in a sea of fine details" (1993, 86). In other words, we don't have recognitional concepts which are as fine-grained as the particular pitches, because we can't remember such fine-grained features. Thus, according to Raffman, nuance ineffability (and so musical ineffability) is the result of memory limitations (1993, 141). As we saw in Chapter Three, recognitional concepts require being able to subsume the feature in question under a category that is stored in memory. You can recognize the black-capped chickadee *as* a black capped chickadee only if you have the appropriate category. In other words, to identify a black-capped chickadee as such, one must apply a category you already possess, a category you retrieve from memory.

According to Raffman, it is in principle possible to name the determinate pitches and intervals; musical ineffability does not come from a deficiency in language, but from "the psychological impossibility of *applying* any such terms 'by introspection'" (1993, 140). We can't apply these terms because we can't remember

which determinate pitch the term applies to. Why can't we remember them? I can think of one initially plausible reason: we just don't have the capacity to store all these different sounds. But there are empirical reasons to doubt that we don't have the mental capacity to store all these sounds. Take the case of a person who has a perfect autobiographical memory.[20] Someone with a perfect autobiographical memory can recall in precise detail experiences they had decades ago, including any sounds. But there is no reason to think that they are any better at identifying determinate pitches and intervals; there is no evidence, for example, that they hear music atonally. For this reason, I doubt that it's the memory *capacity* which is the real source of the problem; the person with perfect autobiographical memory has stored all the sounds she's ever heard, yet still doesn't have terms for the different pitches.

There are two other, perhaps more interesting, reasons why we might have trouble forming the relevant type concepts. The first has to do with our cognitive architecture. As Raffman points out, the reason we have schemas (or concepts) is to reduce the information load. The world does not present itself to us as a Jamesian "blooming, buzzing confusion," because, on the personal level, we are not constantly bombarded by all the information which comes through our five senses. If the whole point of having schemas is to reduce information loads, then we shouldn't expect to have first-person concepts which apply to the early, informationally rich representations. These early representations might be consciously accessible, but they won't be reportable.

A second reason for our inability to form the relevant type concepts is the complex nature of the input. Raffman intriguingly suggests that the problem of forming a concept comes from the variability in the phenomenology of a pitch. She states that the just noticeable difference for pitch varies with "the frequency, intensity,

[20] See Parker et al. (2006) for discussion of such a case.

length, and waveform of the tones heard, as well as with the age alertness, and health of the listener, among other factors" (1993, 85). The sheer number of factors that contribute to how the pitch sounds at a particular time would make it difficult for an individual to form an appropriate long-term representation. While this suggestion is offered as an explanation of nuance ineffability, it also seems to me to be relevant to feeling ineffability, since, if I understand her suggestion correctly, the complexity is phenomenological, rather than representational.[21]

This leads me to the final point in the discussion of Raffman's theory. Nuance ineffability is a feature of particular performances (rather than of the musical work), since the nuances change from performance to performance. The fact that the nuances attach to a particular performance leads Raffman to claim that ineffable musical knowledge is essentially occurrent knowledge: musical performances, unlike musical scores, "are knowable *only in sensing* or *in feeling*" (1993, 94). This feature of nuance ineffability is also a defining feature of feeling ineffability; so, again, it seems that her theory of nuance ineffability might have implications for an account of feeling ineffability.

To summarize, Raffman distinguishes three kinds of ineffability—structural, feeling, and nuance. My interests lie with the latter two kinds of ineffability, and the relationship which might hold between them. Raffman argues that nuance ineffability is the result of the *kind of representations* we use when hearing a piece of music and facts about *how we process* information. The implications of Raffman's account of musical ineffability for an account of the ineffability of visual experience will be explored in section 2.3, after I introduce the perspectival theory of ineffability.

[21] By calling the complexity "phenomenological" I mean merely that what the pitch sounds like is the distinguishing factor—the "just noticeable difference"—between different pitches. I do not want to exclude the possibility that this phenomenology might ultimately be reduced to representational content, as many philosophers suggest.

4.2.2 Lycan's Perspectival Account

The second theory of ineffability I'll discuss is William Lycan's (1996, 2003). For Raffman, the ineffability of music lies in empirical facts about the kinds of representations we can form of musical nuances. Lycan believes that the ineffability of our mental states is a result of the introspective perspective we take on our experiences. Usefully for my purposes, he discusses this ineffability in the context of the knowledge argument. In this section I lay out his position, and in 2.3 I argue that Lycan's introspective account of ineffability collapses into a representative account like Raffman's.

Lycan endorses a view of consciousness which is usually called a "higher-order perception" theory.[22] Higher-order perception theories have long historical roots; they have their origins in Locke's view of consciousness as "inner sense." We discussed Locke's inner sense theory in Chapter Two, and paid particular attention to its relationship to the transparency thesis, the claim that all and only conscious states are states which the subject is aware of having. Contemporary higher-order theorists agree with Locke that in order for a mental state to count as a conscious state some internal faculty must be directed toward that mental state. As Lycan explains, "consciousness is the functioning of *internal attention* mechanisms directed at lower-order psychological states and events" (1996, 14). [23]

---

[22] Other higher-order theorists include Armstrong (1968), Rosenthal (1997). Rosenthal thinks that it is a higher-order *thought*, rather than perception-like representation, which distinguishes between a state being conscious or unconscious, but the views are more alike than different. The higher-order theorists are not without their critics. See for example Shoemaker (1994b, 1994c) for a critique of the notion that introspection is similar to perception. Since, as I argue below, I don't think that Lycan's account of ineffability explains what is going on in the knowledge argument, I'm going to ignore the larger problems facing higher-order theories.

[23] I follow Lycan in using both the words "internal attention mechanisms" and the more commonplace "introspection". Internal attention mechanisms is a broader category that introspection. Lycan believes that consciousness comes in degrees; even animals which did not have full-blown introspective capacities might still have internal attention mechanisms and thus be conscious. See his 1996, pp. 39-40.

One virtue of an inner sense theory, Lycan claims, is its ability to explain the "subjectivity of the mental," which Lycan understands to be the issue raised by the following question: why does it seem from the subject's point of view that her experiences are intrinsically subjective, i.e. cannot be adequately accounted for by a third person description?[24] In other words, in explaining the subjectivity of the mental in a physicalistically acceptable fashion, Lycan has a response to the knowledge argument.[25]

The subjectivity of the mental is the result of the nature of introspective representation. Lycan conceives of the internal attention mechanisms as similar to perceptual processes in many (but not all) ways; one way the inner sense modality is similar to the perceptual modalities is that both present the lower-order psychological state or event from a particular, modality-specific point of view: "Introspection type-classifies mental states—that is, it classifies the various states it surveys into states of the same or different kinds—in its own distinctive way, under its own representations of them" (2003, 391). These "modes of representation" as Lycan sometimes characterizes them, are psychological states, functionally defined by their inferential role in a subject's mind. Most importantly, for my purposes, the introspective

[24] Lycan emphasizes that his inner-sense theory is not meant to resolve any problems about qualia or phenomenal character; rather, the theory is meant merely to distinguish conscious states from non- or pre- or sub-conscious states. On Lycan's account, conscious states are those that we are aware of having because they are (re)represented by internal attention mechanisms.

[25] The first reason Lycan gives for the subjectivity of the mental is the empirical fact that our perceptual systems are processors and filters of information, and the information we receive will depend upon where, precisely, we are situated in our environment. Imagine that you and I are together in a room with a chair. If you are on one side of the chair and I am on the other, then we will have different perceptions of the chair, which will in turn result in our having different "functional profiles" which change in different ways as we move about the room. Assume, for example, we both want to sit in the chair. To satisfy this desire, we would both need to move in different ways. So both the perceptual inputs and the behavioral outputs will be slightly different for each of us, which in turn means that our respective desires will be slightly different. From facts about our perceptual systems Lycan draws two conclusions: that two people will receive different information about the same external environment and that this information will generate different second-order dispositions (1996, 554-55). But this is only part of his story.

representations are not publicly available, because the introspective representations are not synonymous with English or any other natural language.

To explain why these representative expressions are not publicly available, Lycan draws our attention to the perspectival character of pronomial expressions ("my left foot," "I am in Saint Cloud"). A token of the expression "my left foot" has (basically) the same functional role for you and me, but very different extensions. In the case of *de se* expressions, like "my left foot," the distinctive functional role and the particular referent picked out when the expression is uttered nowhere else coincide. You can use this expression to pick out your left foot, or you can use another expression to pick out my left foot, but only I can use "my left foot" to refer to my left foot (1996, 54-62).

Lycan attributes the perspectival character of *de se* expressions to the nature of representation generally, rather than attributing it to a feature peculiar to language. And because it is the result of the nature of representation, the same explanation can be used to explain, in a materialistically acceptable way, why our introspective representations have a perspectival character. The perspectival character of these representations will also make them private. Here is Lycan's own presentation of the explanation for why an individual's mental states are publicly inaccessible:

> If a subject S hosts such a representation [of the subject's own first-order mental states], no one else can use a syntactically similar representation to represent the very first-order state token (of S's own) that is the object of S's own representation. Other people may be able to form syntactically similar representations, but the objects of those representations will be first-order states of their own hosts, not any states of S's. (1996, 59-60; see also 67)

Let me illustrate this with another example of a de se expression. The content of the first-order state, *that my legs are crossed*, is a representation that is unique to me. You can form a similar representation, but when you form it the object is your own first-

order state, which represents your own legs. We can also both use the concept *legs being crossed* as a concept which applies both to you and to me. But when we ascribe the property to ourselves, on the basis of a proprioceptive representation, it follows that the *combination* of the proprioceptive representation and the object of the representation will be unique. You might have a similar proprioceptive representation (these are defined functionally, remember) as I do; and you can refer to the same things that I can; but you cannot have a similar representational state and refer to the same thing.

This analogy is supposed to explain the ineffability of our mental states. On Lycan's view, one major difference between the semantics of expressions like "my left foot" and the internal representation *my left foot* is that the content of the latter cannot be captured by any natural language. Thus the similarity discussed above is *merely* similarity: you and I cannot have an identical proprioceptive representation of our legs being crossed. (This claim should be distinguished from the uncontroversial claim that you and I cannot share the same token representation; Lycan's claim is that the functional state types will be unique for each individual.) This is where Lycan's commitment to a higher-order perception theory comes it. Lycan thinks that when someone is aware of a first-order state, she immediately "tokens a mental word for the type of first-order state being scanned" (1996, 60). This word, according to Lycan, is a semantically primitive private name, a name that it inexpressible in any natural language.

There are two distinct claims Lycan appears to be making. First, that our human language of thought is not English or any other natural language. Second, that each individual's language of thought is private and unique to that individual. The first claim would explain the ineffability of perceptual experiences on its own. If the content of our mental states is not linguistic in nature, then it seems to follow that

there would be some difficulty in adequately expressing the content of our states in English. But this type of ineffability is rather weak; thus far, there is no bar to us finding a way to translate the language of thought into English, especially if, *contra* Lycan's second claim, the language of thought is shared intersubjectively.

If everyone shares a language of thought which individuates our mental states, then we would have in common both the language of thought and the natural language; all we would need to do is to connect the two "languages" in the same way. This might be difficult, but it is not impossible, and, in fact, it seems like something that we actually do. We assume that when others stub their toe they enjoy the same (or at least very similar) mental state as we do when we stub our toe. We can communicate about what the experience is like: the pain comes on suddenly and sharply, and surprisingly strongly, but that it quickly dulls and is gone. To take another example, consider the amount of information which serious wine drinkers can communicate about their experience of tasting the wine. A vocabulary has been formed which connects properties of the experience (i.e. the language of thought representation) of drinking wine, to the words in the English language.

The ineffability which results from the difficulty in translation, which I'll call *translation ineffability*, while real, is a weak notion. With careful attention and training, it seems possible to overcome this form of ineffability; it is similar to Raffman's structural ineffability. However, combined with the second claim that our languages of thought are private, Lycan's ineffability gets some teeth. For we might think that if English and other natural languages are essentially public, and our language of thought is private and unique, then there will be more than mere difficulty in expressing the contents of our mental states in English. It might appear that the translation from language of thought to English is impossible.

Whether the claim that the translation is impossible is true, will depend on how we are to understand the privacy of our language of thought. We can distinguish between a strong and weak notion of privacy. A strong notion of privacy would mean that the privacy of our mental states is a metaphysically essential property of subjective states; that is, even with the most sophisticated futuristic equipment we would be unable to share another's mental state "from the inside."[26] If our mental states were private in this strong sense, then they would certainly be ineffable. Moreover, this ineffability would be metaphysically necessary. If it is impossible for someone else to access to our mental states, and natural languages are essentially public, then it would follow that our mental states are necessarily ineffable.

Lycan does not understand the privacy of the mental in this way. For Lycan, the claim that our languages of thought are private is *not* equivalent to the claim that no one but the subject could have access to her thoughts; his notion of privacy is much weaker. Lycan (following Armstrong) thinks that the privacy of our mental states is merely a contingent fact (2003, 393). It is in principle possible that one day we will be able to introspect the mental states of others, by appropriately wiring up our internal attention mechanisms to someone else's first-order states.[27] So for Lycan, it is just a matter of fact that our mental states are accessible only to the subject. This notion of privacy leads to translation ineffability; it's very difficult for us to connect features of our unique, private experiences with expressions in a common, public language.

I don't want to quibble over terminology, so if Lycan wants to call the above sort of content ineffable, so be it. But it is a weak sort of ineffability and I don't think that even Lycan thinks that this is the whole story. (More on this just below.) The

[26] I have in mind here the kind of private language which troubled Wittgenstein.
[27] If Lycan is right, then with such technology it would also be possible, I would think, for us to hook up our own internal attention mechanisms to what are now unconscious or sub-conscious or pre-conscious first-order mental states.

indexical pronouns give us an example of an intrinsically perspectival entity, i.e. something that cannot be reduced to a third person perspective, which is not threatening to materialism. It's natural to think that this would be a useful model for thinking about the intrinsically perspectival and ineffable nature of mental states.[28] I agree with Lycan that similar puzzles arise in conjunction with the indexical pronouns and the subjectivity of the mental, in particular the irreducibility to a third-person perspective. But I disagree that this is the whole story of why our perceptual experiences are ineffable.

On Lycan's proposal, the ineffability of our mental states results from our having a unique introspective perspective of them. On Lycan's inner sense view, a state S is conscious if and only if S is represented by an internal attention monitor, i.e. introspected. When a state in introspected, it is represented in a language of thought, which is distinct from any natural language. Since the functional concepts that compose this language of thought are unique to an individual, the content of the introspected states in ineffable. But, Lycan argues, we shouldn't draw any ontological conclusions from this ineffability, as we see when we compare Mary's new perspectival knowledge to de se expressions. De se expressions provide a model of an entity that involves a unique combination of functional role and referent.

4.2.3 Comment on Lycan and Raffman

In this section, I argue that Lycan's real explanation for why perceptual experiences are ineffable depends, like Raffman's, on the representations involved, not the perspective. That is, I argue that a complete explanation of the ineffability of

[28] Like some of the philosophers discussed in Chapter One, Lycan connects the subjectivity of the mental to both the problem of understanding the nature of phenomenal character and to the explanatory gap. He thinks that his explanation of the ineffability of our mental states offers materialist solutions to both these problems.

perceptual experiences can't be provided by the perspectival account, but must also include a theory about the nature of the representations. I then argue that while a representationalist theory might explain some of the difficulties we have in linguistically expressing the content of our experiences, there is more to be said. In particular, the representational theory cannot account for what Raffman called the feeling ineffability of our experiences.

What really seems to be doing the work on Lycan's account is the fact that the introspective perspective on our first-order states is similar to perception, and like perception, it delivers its results under a particular, sensory mode of presentation (2003, p. 392).[29] Since this introspective mode of presentation of a perceptual experience is not available to Mary before she leaves her room, she cannot know what seeing red is like.

One of the reasons I have chosen to discuss Lycan's view is that he is upfront about the difficulties facing materialism. He believes, for example, that the phenomenal character of some of our mental states is: "real, internal, specially accessible to/by me, ineffable, intrinsically perspectival, in one sense, inaccessible to science, in one sense, inexplicable" (1996, 11). His higher-order theory not only accommodates these facts about our mental states, but, Lycan claims, predicts them. Especially important for my purposes, he takes seriously the notion that our perceptual experiences of colors are ineffable, as we see in the following imaginary exchange:

> Now suppose a psychologist asks you, "How, exactly, does that patch look to you in regard to color?" You respond, "It looks green." "Yes," says your questioner, "but can you tell me what it's like for the patch to look 'green' to you?" "Um, it looks the same color as that," you say

[29] That it is the quasi-perceptual nature of introspection that is doing part of the work is made evident at the end of his 2003, when he wonders if Rosenthal's higher order *thought* theory would be able to resolve the knowledge argument in the same way. Lycan claims that this is "a further question" (but that he believes it would work since he also believes that intentional states also represent their objects under a mode of presentation). The fact that there is a further question about higher order thought theories indicates that his perceptual model is relevant to the explanation.

> pointing to an issue of *Linguistics and Philosophy*. "No, I mean can you tell me what it's like intrinsically, not comparatively or otherwise relationally?" "Duhhhh." In one way, you are able to describe the phenomenal color paradigmatically as "green." But when asked what it's like to experience that green, if you are like most people, you go tongue-tied. (1996, 124; see also 2003, 6).

Now this example seems to me to illustrate an ineffability which goes beyond the perspectival ineffability described above. This is an ineffability which seems peculiar to perceptual (or experiential) states; there is a reason why Lycan uses color as an example. And this brings us to the main problem with Lycan's account of ineffability. Lycan's higher order perception theory does not distinguish between cognitive and perceptual states; *all* mental states are going to be ineffable on his theory. For the content of all our mental states, cognitive and perceptual alike, will be formulated in the language of thought, which, as we've seen, Lycan believes is distinct from any natural language and is unique to the individual subject. The ineffability that characterizes our perceptual experiences will also characterize our beliefs.

Now Lycan takes pains to point out that his higher-order perception theory does not address the problems associated with phenomenal character. His theory doesn't make a distinction between those states which are characterized by their phenomenal character, and those states which aren't, but between those states that are conscious and those that are not. He thinks that the difficulties phenomenal character poses to the materialist are independent of the nature of conscious awareness which his higher-order theory is meant to explain (1996, 15).[30]

Granting Lycan the claim that the problems of conscious awareness and phenomenal character are mutually independent makes the deficiency of Lycan's account of ineffability stark. Lycan's account of the ineffability of our mental states

[30] Lycan thinks the difficulties associated with phenomenal character can be resolved by intentionalism, the view that phenomenal character is fundamentally representational. Dividing and conquering is a tried and true strategy, and I don't fault Lycan for using it. But I think that he has misplaced on which side of the divide the problem of ineffability lies.

comes from his higher-order theory of conscious awareness; in other words, his account of ineffability has nothing to do with the phenomenal character of the mental states. But we don't think that the ineffability of our beliefs (granting that there is such a thing) is of the same kind as the ineffability of our perceptual experiences. Lycan's explanation fails to explain this deeper sense of ineffability, which seems most prominent in our perceptual experiences.[31]

To fully explain the ineffability of our perceptual experiences, there needs to be a way to distinguish the cognitive states from the perceptual states.[32] In keeping with the spirit of his higher-order account, here are two suggestions for where to locate the distinction between perceptual and cognitive states. First, perceptual and cognitive states could be distinguished by the type of internal mechanisms that take these states as inputs. On Lycan's account, there are multiple internal attention mechanisms, and it could be that different attention mechanisms represent differently. Perhaps there is a division of labor, and one type of internal attention mechanism is directed toward first-order perceptual states and a different type is directed toward cognitive states; then each type of attention mechanism generates a distinct kind of second-order representation. A second place to locate the distinction is at the lower level; it might be that the nature of the first-order representations of perceptual experiences and first-order representations of beliefs are different. In other words, the representations fed into the attention mechanisms are of different sorts from the beginning.

Ultimately, which, if either, option is correct is an empirical question. I bring up these two options in order to point out that in both cases the distinction between cognitive and perceptual states ultimately depends on differences in kinds of content.

---

[31] Pitt (2004) argues that we should take into consideration the phenomenology of belief.

[32] One way to do this, which is discussed below, is to use the notion of nonconceptual content. According to the nonconceptualist, perceptual states are distinct from belief states because they are not composed of concepts.

The question of ineffability is a feature of phenomenal character and not a feature of the conscious awareness of the perceptual state. In other words, it is not the higher-order perspective on our experience which is responsible for the ineffability; it is a feature of the experience itself.

This brings us to Raffman's claim that the ineffability of musical knowledge is due to empirical facts about how we represent musical performances. In a way, her theory suffers from the opposite problem of Lycan's: Raffman's theory is proposed on the basis of highly specific claims about how we hear and understand tonal music. But I think that if her account is correct, then there is good reason to believe that similar accounts are available for the other senses. So, let's assume that a representational theory of ineffability something like Raffman's is correct, and apply it to Mary's situation. (To avoid (mis)attributing the view to Raffman, in what follows I speak generically of "the representationalist theorist.")

In the role of pitch will be determinate colors. According to the representationalist theorist, typical human subjects will have no words for picking out specific determinate colors, like yellow108. On the representationalist theory, this is the result of two features of cognition. First, there is the representation itself. Like the representations we have of pitch, we might assume that our representations of determinate colors are complex and variable. There are variations both across subjects and for the same subject at different times. Take lighting conditions for example. It's highly unlikely that across different viewings of a determinate shade the lighting conditions will be the same. Thus each time we see yellow 108, it will appear slightly different to us, and thus be difficult to classify.

The variability in the representations of a determinate shade will make it very difficult to classify as a type of shade, but it's still possible that with enough concentration and effort one could manage to recall the last few times you saw

yellow108 or a similar shade and compare and contract the appearances in order to make an identification. Of course, this is not how ordinary human subjects proceed. And this is because of the second relevant feature of human cognition: the memory constraint. It turns out that we cannot imaginatively contrast and compare similar shades of yellow because we can't recall the determinate shades.

Both of these features – the variability in representations of pitch and the constraints on memory – are just part of the cognitive architecture for human beings. At the lower levels of representation, lots of information is encoded; as you move upstairs, some of this information is lost. Our memory works on representations that are on the upper levels; we don't store all the perceptual information we receive. All of this seems perfectly reasonable, and goes some way to explaining why we don't have words for the determinate color shades. We can't remember shade representations, and thus we can't subsume our experience of a shade under a concept.

On the representationalist view, the ineffability of perceptual experience is a contingent fact about our cognitive make-up. The representationalist theory explains why people are tongue-tied when it comes to explaining what their experiences are like. But the ineffability that is explained by the representational theory is obviously of the weak variety. The representational theory does not explain the more basic issue of why having the experiences is necessary for acquiring the concepts in the first place. In the last section of the chapter, I will offer a proposal based on the phenomenology of visual experiences that goes some way towards explaining why having the experience is required for having the concepts.

4.3. Ineffability and Nonconceptual Content

Philosophers do not often discuss ineffability directly, but a closely related subject, the possibility that the content of perceptual experience is nonconceptual, has

received much attention in recent years.[33] Those who believe that the content of perceptual experience is nonconceptual - the *nonconceptualists* - argue that the content of experience is significantly different than the content of belief (the paradigmatic conceptual mental state); that is, the nonconceptualists argue that perceptual and conceptual states have fundamentally different kinds of content.

As we'll see, there is some confusion about what, precisely, the nonconceptualists are claiming. But first I want to introduce an argument which shows that if the content of visual experiences is nonconceptual, it is ineffable. Before we begin, let me make a brief, but important, terminological point. The literature on nonconceptual content speaks broadly of the "content of perceptual experiences," which might include both phenomenal and nonphenomenal beliefs. Take, for example, a visual experience of seeing a red apple on the desk. Much of the content of the experience can be captured by the belief that there is a red apple on my desk, a belief which can be nonphenomenal, in the sense described at the beginning of this chapter. (Someone outside the room can come to believe that there is a red apple on my desk on the basis of my testimony.) In what follows, I will speak of content as being "partly ineffable" to allow for the possibility that our experiences that much of the content of our experiences is expressible.

Richard Heck has drawn a connection between a perceptual content's being nonconceptual and its being inexpressible. After defending nonconceptualism, he claims it follows that:

> …there is a certain sort of gap between the reasons we have for our beliefs and the reasons we can communicate to others: If someone asks me why I believe, say, that there is a desk in front of me, I can do no better than to say that it appears to me that there is a desk in front of me (or, less formally, that I see it there); the fact that it does so appear to me, that is, my being in a certain

[33] See Gunther (2003) for a collection of classic and contemporary papers on nonconceptual content and a helpful Introduction.

> perceptual state, is what gives me a reason for my belief, but *I cannot tell you (or myself, for that matter) exactly what content my perceptual state has (that is, exactly how the world does appear to me)*. (2000, pp. 519-520; italics mine)

I take it that the italicized part of the above passage is equivalent to the claim that the content of perceptual experience is partly ineffable. The difficulty, as Heck notes when he points out that we cannot even articulate to ourselves the content of our experiences, is not merely that a person must have an experience before she can form beliefs with the relevant phenomenological content. The difficulty runs deeper; according to Heck, while undergoing an experience we are unable to express its phenomenological content, even to ourselves.

Here is a more formal argument for the claim that nonconceptual content implies ineffability:

(1) If a content is entirely linguistically expressible, then it is conceptual.

(2) The content of perceptual experience is not conceptual.

(3) Therefore, the content of perceptual experience is not entirely linguistically expressible, i.e. it is partly ineffable.

Premise (1) seems pretty plausible. If someone is able to linguistically express the content of a thought, we take that as evidence that she has the corresponding concepts. And while there is much debate about what these concepts are, most people agree that there are lexical concepts that correspond to words in natural languages. Since most philosophers would be inclined to accept (1), I am going to assume that if the content of perceptual experiences is nonconceptual, then it is ineffable. But this doesn't take us too far. Premise (2) still needs to be defended.

The rest of the paper looks at three arguments, each based on a different phenomenological feature of visual experiences, in favor of premise (2). There are two reasons for examining these three arguments. First, as discussed above, nonconceptualism seems to imply ineffability. This implication might just follow

from the claim that the content of perceptual experiences fundamentally differs from the content of cognitive states. If the nonconceptualist is correct that these perceptual experiences and cognitive states have different kinds of content, then it is natural to think that this might have implications for how we acquire knowledge about the objects these states are about. That is, if the content is different in kind, then we might naturally expect that there would be difficulties transforming the content of the experiential states into the kind of content had by cognitive states. Or, if a necessary condition on linguistic expressibility were that a content is conceptual (i.e. premise (1) above), then this would also explain why nonconceptualism leads to the ineffability of perceptual experiences.

But there is another reason for looking at the following three arguments. A close examination of the phenomenological claims underlying the arguments hasn't been done, and the three features are sometimes treated in the literature as more or less interchangeable in arguments for nonconceptualism.[34] As we'll see, once the phenomenological claims are fleshed out, two of the arguments fail, and the third is inconclusive. However, in the course of examining the phenomenology of visual experience we learn that the properties investigated are of two different kinds; the first two properties discussed are contingent, and the third is necessary. This is an interesting discovery in itself. But it also suggests that this third property—determinacy—is responsible for the strong ineffability of visual experience.

4.3.1 The Argument from Richness

The three arguments for nonconceptual content I want to look at are the Arguments from Richness, Fineness of Grain, and Determinacy.[35] While versions of

[34] So, for example, discussions will take "richness" to be shorthand for "richness of detail" which confuses two of the features.
[35] See Byrne (2004), Evans (1982), Kelly (2001b), Heck (2000), McDowell (1994), Peacocke (1992), Speaks (2005), Tye (2006) for discussion of these types of arguments.

these arguments can be found in many places, the three phenomenological features are frequently treated together. This is a mistake, for the three features are independent of one another.

Here is a brief characterization of each feature; they are each discussed in more detail below. A phenomenological content is *rich* if it carries a lot of information. Our ordinary visual experiences are rich, but in atypical settings, like a psychology lab, they might not be. A phenomenological content is *fine-grained* if the information it carries is presented in a way that allows for small, phenomenological differences. For example, the phenomenological content of visual experiences for humans with ordinary vision allows for subtle discriminations between color shades. Finally, a phenomenological content is *determinate* if it presents the world as being unambiguously a certain way. When we look at a green object, for example, the object appears to be a specific shade of green; a green object does not appear as both lime green and forest green simultaneously.

The Argument from Richness begins with the claim that most of our perceptual experiences provide us with more information than we can adequately characterize. So, for example, I'm currently perceiving the screen of a laptop with text on the left and icons on the right, a red computer sleeve sits a bit behind the computer and to the right, in front of that is a water bottle next to a yellow pencil, etc. At any given moment, my visual experience presents me with more detail than I could possibly grasp; so, the Argument from Richness goes, its content must be nonconceptual.

This argument moves pretty quickly. While it is certainly true that most of our experiences are rich in this way, in that they have more information than we generally bother to acknowledge, it's not clear that it follows that the content of experience is nonconceptual. The Argument from Richness seems to be this:

(1) I have a visual experience with a rich content R .
(2) If R were composed of concepts, $C_1 \ldots C_n$, then I would possess $C_1 \ldots C_n$ (since I am enjoying an experience with content R).
(3) I do not possess $C_1 \ldots C_n$.
(4) Therefore, R is not composed of concepts, $C_1 \ldots C_n$.

Let me first offer some remarks of clarification. Premise (1) is meant to establish that the case is one in which the visual experience represents many objects and properties. So, for example, you should imagine a typical visual experience (like the one described above or the one you are currently having) and not the type a subject might have in the psychology laboratory. This much should be uncontroversial; most of our visual experiences are rich in this way.

There are, however, many ways to understand what it means to say that a visual experience is rich.[36] "Richness," as I am using the word, is a quantitative predicate. If the content of an experience is rich, then it presents a large number of objects and properties. It's difficult to be much more precise about what counts as "a large number." Richness seems like a comparative property, but it's hard to determine what should count as the comparison class without undermining the above argument. If richness is to be used in an argument for nonconceptual content, then it can't be that visual experiences are rich relative to the concepts we possess, because this would beg the question against the conceptualist. If richness is used to support ineffability, then it can't be that visual experiences are rich relative to the number of words we have for describing them without begging the question against someone who denies that visual experiences are ineffable. But this vague formulation will do for our purposes.

In one of the earliest arguments for nonconceptual content, Fred Dretske brought to the attention of philosophers a series of experiments by the psychologist G.

[36] I am interested in only the information that is made available by the phenomenological content of the experience, so I'm not concerned with any information at the subpersonal level that the experience might carry, i.e. information that is not available to the subject of the experience like that carried by early stages of visual processing.

Sperling.[37] These experiments illustrate one relevant way an experience can be rich. Sperling showed subjects an array of nine letters for a brief period of time (50 ms). After this brief exposure subjects reported a persisting visual image, and generally could recall three or four of the letters in the array. To figure out what information the subject could access, Sperling introduced another element into the experiment. Right after the array was removed (150 ms) a marker would appear indicating one of the rows or columns in the array. The subjects accurately identified the letters in the marked row or column. Since the subject had no idea which column or row would be marked, it seems that all of the information in the array must have been available to her when the marker appeared.

The information available to the subjects was presented phenomenologically. This is supported both by the subjects' own claims of experiencing a persistent visual image, and by an additional variable Sperling added to the experiments. During the trials, Sperling varied the brightness of the background against which the array was shown. As would be expected if the information were presented visually, when the background was bright, the subjects' success rates were significantly worse than when the background was not so brightly lit. (Imagine looking at a PowerPoint presentation with the lights on.)

Sperling's experiments show three relevant things about visual experiences: first, they are informationally rich (we do have information about all the letters in the array), second, this information is presented phenomenologically to the subject, and third, access to this information is limited to a short time interval. Dretske uses these facts to argue for a distinction between the way perceptual and cognitive systems encode information; he argues that the perceptual systems are not subject to the same limitations as the cognitive system, a fact he attributes to a difference in how the two

[37] Dretske (1981). An excerpt is reprinted in Gunther (2003). See also Tye (2006).

systems encode information. Dretske's position is a clear example of nonconceptualism; he argues that the content of beliefs and the content of perceptual experiences are different *in kind* because they encode information in fundamentally different ways.

But there is another way to interpret Sperling's results. It is possible that the problem the subjects face in describing their visual experience is computational; the subjects don't, for example, have time to form all the relevant beliefs before their memory of the array fades, which would affect their ability to describe their experience. From this computational limitation, it doesn't follow that the content of the experience is different in kind from the cognitive content.[38] All that follows from the Sperling experiment is that a subject can enjoy a perceptual state without being in a position to deploy concepts for all the objects and properties represented by the visual state.

Even if this argument encounters problems as an argument for nonconceptual content, it still might shed some light on why the content of perceptual experience is ineffable. As Sperling's experiments show, our memories of perceptual experiences deteriorate rapidly; if experiences are this fleeting, then it's no wonder that we are unable to completely articulate them.

The fleetingness of memory adds much, I think, to the feeling that experiences are ineffable. This feature is more vivid in the case of tastes and smells; imagine, for example, the taste of a pungent cheese. It's hard to describe much after the first bite; the initial flavor seems to recede as one tries to examine it. Intense bodily sensations and emotions are also ephemeral in this way. It feels that if it were possible to pin

[38] Tye says, "Sperling's experiment supports richness, but visual experiences could be rich, as revealed in that experiment, without having a nonconceptual content. For the thesis of richness alone does not rule out the possibility that visual experiences are conceptual states whose conceptual contents contain more information than the belief-forming processes can handle under certain constrained circumstances" (2006, 519).

down the feeling, if it would stay still, then we would be able to express it in words. But our experiences are not like snapshots, not even our visual experiences. We are constantly moving around our environment, and that environment is constantly changing. In typical circumstances, we don't have enough time to describe our experiences before they (and our memory traces of them) are gone.

While the fact that our experiences are fleeting goes some way to explaining the difficulty we have in expressing them linguistically, this would support only the computational form of ineffability. If their fleeting nature were the reason why we couldn't verbally express our experiences, because articulation of them takes time, then this would be due to a computational limitation on our short-term memories. The fleetingness of our experiences, which is made explicit by Sperling's experiments, does not provide reason for thinking that the ineffability of visual experience is of the strong kind.

4.3.2 The Argument from Fineness of Grain

Our visual experiences represent the world in minutely detailed ways. So, for example, when I look at the pencil lying on my desk, the yellow on the top appears brighter than the yellow on the sides; the green metal band at the eraser end shines where the light strikes it; there are numerous tiny dents in the wood; the texture of the eraser is rough at the end, but smooth around the sides, etc. But I don't have concepts for the particular shape of the dent in the wood, or even for the particular shade of yellow; thus, the Argument from Fineness of Grain goes, the content of perceptual experiences must be nonconceptual.

The Argument from Fineness of Grain is structured like the Argument from Richness. The first premise states what seems to be an obvious fact about the content

of a visual experience, and the second premise points to a limitation of the subject's concepts.

(1) My visual experience is fine-grained.
(2) If my visual experience were conceptual, then I'd have the relevant fine-grained concepts.
(3) I don't have the relevant fine-grained concepts.
(4) Therefore, my visual experience is nonconceptual.

Much of the debate between the conceptualists and the nonconceptualists has focused on the third premise. When faced with an argument like this one, conceptualists like John McDowell deny premise (3) on the grounds that subjects don't need a special concept for (say) each shade of color.[39] Instead, we can think of properties presented in our experiences demonstratively: I can think to myself *that shade* while attending to the yellowness of the pencil, or *that shape* when focusing on its shape. Since these demonstratives exhibit the right degree of fineness of grain, and the ordinary human perceiver possesses these demonstratives, premise (3) is false.

McDowell is correct that if we recognize that demonstrative concepts can be constituents of conceptual content, then (3) is false. Because they are tailor-made to the properties they pick out, demonstratives will be relevantly fine-grained. So the above argument doesn't work.

It seems to me, though, that McDowell's response is ultimately misguided, because thinking to myself *that shade* while looking at the pencil is an application of demonstrative concepts to the experience; the experience itself does not present these properties demonstratively. As others have pointed out, the availability of demonstratives doesn't show that experience is conceptual, but rather that we can apply concepts to all properties represented by our experiences.[40] The availability of

[39] See McDowell (1994), pp. 56-59. A different way to phrase McDowell's objection is that he is denying the assumption which seems to underlie premise (2), the assumption that the "relevant fine-grained concepts" are ordinary predicates.
[40] See Heck (2000), Tye (2006).

demonstratives allows us to pick out fine-grained and determinate features of our experience, but the availability of demonstratives doesn't show that the content of experience is itself composed of demonstrative concepts.

It's hard to evaluate claims about whether or not experience presents objects and properties demonstratively. For a more decisive objection to McDowell, the nonconceptualist needs an argument that the fineness of grain of perceptual experiences is different from the fineness of grain of concepts; that is, to respond to McDowell's objection to (3), the nonconceptualist needs an argument that these two instances of "fine-grained" in premises (1) and (3) pick out different properties.

Let's begin with the fine-grained nature of experience. The online edition of the *Oxford English Dictionary* defines "fine-grained" as "consisting in particles that are very small," and many of the relevant *OED* definitions of the word "fine" also involve the idea of being broken down into small pieces. This suggests that one idea behind fineness of grain is that our experiences can be decomposed into very small parts; let's call this the *structural sense* of "fineness of grain." We might speak of a cup being "red", but when we visually attend to the cup we find that its appearance cannot be described by such a simple word. Our visual experiences of cups do not present them as of a uniform color; instead, the appearance of the cup is complicated by factors like the way it is illuminated, and the material it is made of. These complexities in the phenomenological content of our visual experience of the cup are not, arguably, due to sophisticated concepts of the subject. Any adult human with a normal visual and cognitive system would notice these details if he or she attended to them.[41]

[41] Interestingly, this is not so obvious in other cases, e.g. the novice's and the expert's experience of a particular wine.

In contrast to this structural sense, there is a *comparative sense* of "fineness of grain," which is used in discussions of concepts. So, for example, the following list of concepts goes from coarse-grained to fine-grained: ANIMAL, BIRD, SPARROW, HOUSE SPARROW. Each concept is more fine-grained than the one that precedes it. And our color concepts exhibit similar degrees of grain: COLORED, RED, DARK RED, CRIMSON. These concepts compose the contents of beliefs, which can also be described as more or less fine-grained; to have the belief that the bird in the tree is a house sparrow rather than the belief that the animal in the tree is a bird is to have a more fine-grained belief.

This comparative sense is different from the structural sense of "fine-grained"; the structural sense makes a claim about the constitution of our visual experiences that does not seem to apply to our beliefs. My belief that the cup is red just doesn't decompose into tiny particles, in the way that my experiences do. So it looks like there is a property that the contents of my perceptual experiences have, but that the contents of my beliefs lack. If this were the case, then it would be the beginning of an argument for nonconceptual content. Unfortunately, I don't think this argument works because the phenomenological sense of "fine-grained" ultimately depends on the abilities which underlie the conceptual sense of "fine-grained."

Let me explain. Almost any account of perceptual concepts is going to identify concepts with recognitional abilities or require recognitional abilities as part of the possession conditions.[42] If someone isn't able to phenomenally discriminate the shades G1 and G2, then their phenomenological content is the same for that person. And if their phenomenological content is the same, then it follows trivially that the shades are not distinguishable on the basis of how they visually appear. In other

[42] For example, in order to attribute to someone the concept DOG they must be able to reliably recognize dogs. If someone cannot reliably recognize dogs when she encounters them, then she fails to possess the concept DOG.

words, the same abilities that are used to individuate phenomenological contents will also be possession conditions for concepts. Any relevant notion of fineness of grain is going to depend on our concepts, which means it cannot be the feature that distinguishes nonconceptual content from conceptual content. So, the nonconceptualist doesn't have a good response to McDowell.

More significantly for my purposes, because the fine-grained nature of our experiences is determined by our abilities to phenomenally discriminate between different types of experiences, and those discriminatory abilities are contingent, it is a contingent feature of the phenomenological content. And a contingent property of our experience can't explain what is supposed to be a necessary property of our experience, namely, that it can't be linguistically expressed. Given that fineness of grain is determined by our discriminatory capacities, if the phenomenological content of visual experiences were ineffable because it is fine-grained, then this ineffability would be merely computational. Thus, fineness of grain is not the right type of property for explaining strong ineffability.

4.3.3 The Argument from Determinacy

In addition to being fine-grained, perceptual experiences are also determinate, by which I mean that they present properties in precise and specific ways. My visual experience does not, for example, present the pencil only as yellow, but as a particular shade of yellow. By presenting properties in this precise and specific way, our experiences present the world to us without ambiguities. When I look at the pencil it seems to me to be this way rather than that way. I might hesitate over how to categorize a certain property of the pencil—is it orangey-yellow or a yellowish orange? --but the experience itself is not ambiguous. And this is true even for people with poor vision. When someone with poor vision takes off her glasses, the world

looks different, but the difference isn't a new lack of determinacy. Instead, the world appears to have different properties, ill-defined boundaries for example. Still, I want to claim, the experience of the person with poor vision is determinate: it presents the world in a particular way, a determinately fuzzy way.

Fineness of grain and determinacy, considered as properties of experiences, are distinct properties. It is easy to imagine having a phenomenological content that is not fine-grained, but still determinate; if we could make fewer distinctions among color shades, for example, this wouldn't affect the determinacy of the color appearances. So we can have determinacy without fineness of grain. But determinacy is arguably an essential property of phenomenological content; I, at least, have a hard time imagining what it would be like to be presented with ambiguous phenomenal properties of objects. For example, when I look at the duck-rabbit drawing I see a picture of a duck or a rabbit, but I don't see both a duck and a rabbit simultaneously; there is no ambiguity about which figure I am seeing. And the phenomenological content of a visual experience of a red paint sample is not ambiguous between different shades of red; the sample is presented as a particular, single shade of red.

Because determinacy seems to be an essential property of phenomenological content, rather than a property that arises from our abilities, this phenomenological feature is better suited than richness or fineness of grain for supporting both nonconceptualism and ineffability. The following argument captures the idea that the content of visual experiences and conceptual content are fundamentally different, in a way that is not dependent on the perceiver:

> (1) The content of visual experience is determinate.
> (2) Conceptual content is not determinate.
> (3) Thus, by Leibniz's Law, the content of visual experience is not conceptual content.

This type of argument is what is needed to support nonconceptualism. But premise (2) is hard to interpret. Without a particular account of conceptual content in mind, it is hard to evaluate, and evaluating competing theories of conceptual content, of which there are many, is beyond the scope of this dissertation. But it seems that many of our concepts are vague, in the sense that it is indeterminate what falls in their extension. It's possible that many of our concepts are like the concept BALD. If this were true, if many of our concepts, and in particular many of our perceptual concepts, were vague, then this line of argument might be decisive, both as an argument for nonconceptualism and an explanation of ineffability.

Conclusion

I have argued that a plausible explanation behind the central intuition of the knowledge argument (the intuition that Mary cannot know certain truths about color perception until she sees colors) is that the phenomenological content of these truths is ineffable. I distinguished between two types of ineffability that might explain why having experiences of colors is necessary for the formation of the relevant beliefs about colors. I argued that if the ineffability of phenomenological content is the result of computational limitations, like memory and processing speed, then we still lack an adequate explanation of the central intuition behind the knowledge argument. While these computational limits help explain why we have difficulties in articulating the phenomenological content of our visual experiences, they don't explain why Mary lacks the resources to even attempt to describe the content of color experiences she has never had.

Thus, if the Ineffability Proposal is to offer an explanation of Mary's predicament, it must be interpreted as a claim about the nature of the phenomenological content itself, rather than as a claim about the abilities of the

subjects of experience. After close examination of the three phenomenological features most frequently used to argue that the content of perceptual experience is nonconceptual – richness, fineness of grain, and determinacy – I argued that only determinacy is an essential property of phenomenological content, and thus the right type of property to explain why experiences are necessarily ineffable.

There is no direct argument from the claim that phenomenological content is determinate to the claim that it is metaphysically ineffable. Any argument from the determinacy of phenomenological content to the conclusion that phenomenological content is metaphysically ineffable will require a premise about conceptual content, a premise that claims conceptual content lacks the kind of determinacy possessed by phenomenological content. Since defense of such a premise lies beyond the scope of this dissertation, a full evaluation of the Ineffability Proposal will depend on further investigations into the nature of conceptual content.

But we've made good progress. While doubt has been cast on the theory that the content of perceptual experience is nonconceptual, there is no reason for thinking that the phenomenological content of visual experiences cannot be both conceptual and metaphysically ineffable. If phenomenological content is understood as shared by both the perceptual and cognitive states that Mary undergoes when she sees her first lemon—the visual experience, the belief about the way the lemon looks, and the belief about what seeing the lemon is like—then it needs to be conceptual. And we have discovered a property of visual experience, determinacy, which combined with a theory of conceptual content, is capable of explaining why the nature of perceptual experience is ineffable.

Chapter Five

Concluding Remarks

This dissertation is an examination of *the central intuition* behind Frank Jackson's knowledge. I explore different ways of understanding what lies behind the common conviction that Mary cannot know everything about colors until she has had a color experience herself. Jackson himself uses the intuition to challenge physicalism, but his reliance on the central intuition brings up an issue which also needs to be addressed by the dualist: *why* can't Mary know what seeing red is like before having the visual experience herself? Dualism isn't threatened by the central intuition in the way that physicalism is, but as a rival theory of mind it owes some explanation for why, assuming the central intuition is sound, certain experiences cannot be understood by a person who has not entertained them. If it turns out that experiences are not physical, this will do nothing to deepen our understanding of why Mary doesn't know what it's like to see red while stuck inside her room. That is to say, even the dualist needs to explain what makes certain mental states, like knowing what it's like to see red, different from other mental states, like knowing that fire trucks are red.

The intuition that having experiences is required for acquiring knowledge has been around a long time; the central intuition is a specific instance of the modern empiricist claim that all knowledge requires experience. In Chapter Two, I looked at the historical roots of the general claim. I pointed out that Locke relies on thought experiments which invoke the central intuition; foreshadowing Jackson's own thought experiment, for example, Locke imagines the case of a boy raised in a black and white

room. Locke uses these thought experiments to support the general empiricist contention that all knowledge comes from experience; that is to say, Locke uses the cases where the central intuition is strongest to make his case for the general principle and doesn't appear to think that the central intuition itself might need support. Because his interest lies in the general empiricist claim, the question that I have just raised – what makes these beliefs about colors different from other sorts of beliefs – isn't raised. In this, Locke is like the other empiricists; he doesn't distinguish between phenomenal and nonphenomenal beliefs.

After discussing the historical background of the central intuition, I turned to what might be called an epistemological explanation of the central intuition. This account, advocated by Lewis, Nemirow, and Mellor, claims that knowing what the experience of seeing red is like belongs to a special category of knowledge, know how. According to this account, know how is fundamentally different from propositional knowledge, primarily because it can only be acquired by having the requisite experience. The know how account is interesting to me for two reasons: first, if true it would provide a satisfying explanation of the central intuition and second, it provides this explanation in part by focusing on the ineffability of our knowledge of color experiences.

According to the know how account, Mary can't know what it's like to see red until she sees red because knowing what it is like to see red consists in the abilities to recognize, remember, and imagine red. Mary lacks all three of these abilities while in her room. Moreover, the only way to gain these three abilities is by actually having the experience of seeing something red. One cannot be told how to recognize, remember, and imagine red; one must have the experience of seeing red in order to gain the abilities. According to Lewis, Nemirow, and Mellor, knowing what it's like to see red should be treated on a par with knowing how to ride a bike or play the

mandolin; no amount of propositional knowledge will be sufficient for acquiring the abilities necessary for being ascribed the knowledge.

The know how response to the knowledge argument explains the ineffability of our phenomenal beliefs by subsuming them under a category of knowledge which is made up of examples of knowledge which clearly do not threaten physicalism. No physicalist, for example, is worried that we can't teach someone how to ride a bike merely by describing the underlying physics. But one needn't be a physicalist to find the know how account appealing. As I mentioned, everyone who intends to offer a complete account of the mind needs some explanation for why the central intuition is true. On the know how account, the central intuition is explained by (i) pointing out that there is an entire class of knowledge which is ineffable yet non-mysterious and (ii) showing that knowing what color experiences are like falls under this category. If the proponents of the know how account had been able to satisfy (i) and (ii), then the central intuition would have been adequately explained. However, I argued that the proponents of the know how account fail in their attempts to address (ii); that is, they fail to show that knowing what it's like to see red consists in the abilities to recognize, remember and imagine red. In particular, the know how account has a problem in accounting for the situation in which someone is reflecting on her occurrent experiences of a particular color. It seems possible that someone might know what it's like to undergo a certain experience even if she lacks the abilities to recognize, remember or imagine it.

I think that the know how account goes wrong by focusing on the epistemological status of Mary's mental states. The fundamental difficulty raised by the thought experiment involves her inability to entertain a belief of a certain kind; whether or not this belief counts as knowledge is a secondary concern. Mary can't know what seeing red is like because she can't entertain the relevant belief. What is

needed to explain the central intuition is not a distinction between kinds of knowledge, but a distinction between kinds of beliefs. In other words, I don't think that the central intuition can be explained by appeal to different ways of knowing; the central intuition is grounded in a metaphysical distinction between types of mental states.

I focus on the ineffability of visual experiences in the final chapter. The ineffability affects both those who try to describe the experience and those who try to understand the description. Those who know what color experiences are like cannot adequately express what it's like to see colors and even if color experiencers could adequately express such knowledge Mary would fail to understand it. I pointed out that the ineffability of our visual experiences might come from either one of two sources, which I've labeled weak and strong ineffability. On the one hand, the ineffability might be the result of computational limitations, e.g. memory, on human abilities to think about visual experiences. On the other hand, the ineffability might be due to some essential feature of the phenomenology of human visual experiences.

I chose to frame the issues raised by the central intuition in terms of ineffability in order to highlight the fact that the central intuition is not just a problem for the physicalist, but an issue that needs to be resolved no matter what one's ontological commitments. Jackson's knowledge argument is interesting not only because of its challenge to physicalism, but because, as the subsequent literature shows, it reveals the strength of the intuition that no amount of linguistic description is going to help Mary understand what seeing red is like. The thought experiment vividly illustrates that, at least for a certain class of truths, it appears that the empiricist thesis was correct. And it seems to me that there is an interesting question of explaining why the empiricist thesis holds true for these truths but not others.

But how the central intuition is explained does bear on the issue of physicalism. If a convincing explanation for the ineffability of our color experiences

can be given, then this will bolster the physicalist case against the knowledge argument, while avoiding the necessity of directly addressing issues about the nature of the physicalist commitment, in particular the debate between the a priori and a posteriori physicalists. If it can be shown that the nature of our experiential mental states is such as to preclude linguistic expressibility, then this would help to vindicate either type of physicalism, though the support it offers is different in each case.

It would help an a posteriori physicalist strategy like Horgan's, discussed in the first chapter, by completing the explanation. Horgan argues that the physicalist is not committed to the claim that every truth about, for example, color experience, is expressible in physicalist language, but is merely committed to the claim that all the terms used in describing color experience refer to physical entities. An explanation of the ineffability of perceptual experiences would complete this argument, by explaining *why* certain beliefs can't be expressed in scientific terms (or any terms at all).

An explanation of the ineffability of perceptual experiences might give the a priori physicalist a reason to question the common conviction in the truth of the central intuition. If it turned out that visual experiences are only weakly ineffable, then we should be wary of assuming that an omniscient being like Mary, who is not subject to the same limitations as normal human subjects, would fail to know what seeing red is like while still in her room.

A second reason for framing the issue raised by the knowledge argument in terms of ineffability is to make room for a different approach to phenomenal concepts. As I mentioned in Chapter One, many contemporary philosophers believe that the knowledge argument appears to pose a problem for physicalism only because philosophers have failed to adequately acknowledge the special nature of phenomenal concepts. Once we recognize the special nature of these concepts, these philosophers claim, we will be in a position to respond both to Jackson's knowledge argument and

Levine's explanatory gap. According to the phenomenal concept theorists, phenomenal concepts are special because, unlike concepts like DOG and RADIO, possessing the phenomenal concepts requires having the requisite experience.

While I agree with the phenomenal concept theorists that possessing these concepts requires having the experience, there seems to me to be a further question about how to understand the nature of experience which might explain why these concepts can only be acquired in this way. The phenomenal concept theorists do not offer much of an explanation; like Locke and the early empiricists, they typically rely on our intuitions about such cases to make their argument. By closely examining the features of experience which are responsible for their ineffability, but trying to steer clear of the difficulties which arise in any discussion about the nature of concepts, I also hope to have shed some light on the nature of these phenomenal concepts and why their acquisition requires experience.

The enduring appeal of Jackson's thought experiment is due in part, I think, to the fact that it focuses our attention on just one single aspect of visual experience. We imagine Mary, upon her release or still in her room, shown a prototypical color specimen—a ripe tomato—rather than imagining her throwing open the door to a colored world. Focusing on a particular color experience, like that of seeing a ripe tomato for the first time, rather than seeing the entire colored world, has consequences for how we understand which notion of ineffability is relevant. I have argued that only ineffability of the strong sort, as a feature of the visual experience itself, will explain the central intuition. Weak ineffability, which is due to cognitive limitations of the human subject and in particular the constraints imposed by memory that make it difficult if not impossible to re-identify fine-grained shades of color, is not an issue for Mary. The problem Mary faces is that she cannot form any beliefs about any shades at all, not even the unique hues (which the typical human can re-identify).

I think that weak ineffability can account for many of the problems most of us face in trying to describe our ordinary, daily experiences. When we search for the right words to describe the color of the wallpaper we liked or the meal we had last night much of the difficulty comes from trying to remember the precise details. But I also think that there is a deeper sort of ineffability at work, an ineffability which is a property of the phenomenology of the experiences themselves. This strong ineffability would explain why we cannot provide Mary with the sort of information that would allow her to think about a single shade of color. Even when we are in the midst of having an experience of a certain color, we can't adequately describe it. And it's not because of her cognitive limitations that she can't form the right sort of thought; in addition to there not being a question of re-identification, by hypothesis Mary isn't subject to the same cognitive limitations as ordinary humans.

The dissertation ends with a brief exploration of three of the phenomenological features of visual experience which might account for its ineffability. Like Berkeley, I focused on phenomenological features which are peculiar to visual experience, but it's obvious to me that ineffability is a feature of all our sense experiences, including our emotional states. For this reason, one might think that by concentrating on phenomenological aspects particular to visual experience that I'm looking in the wrong place; instead, I should be looking for some common feature of human sensory experience which accounts for its ineffability. I think, however, that the determinate nature of our visual experiences is also a feature of other kinds of experiences. The sound of Johnny Cash's voice, the taste of a tomato plucked from the vine, the smell of the lilac bush, the feeling of jumping in a cold river on a hot summer day are all experiences which share this feature. None of these experiences can be adequately expressed in words and knowledge of them isn't available to those who have never had the experience.

REFERENCES

Alter, Torin, and Walters, Sven. 2007. *Phenomenal Concepts and Phenomenal Knowledge.* Cambridge, Mass.: Oxford University Press.

Armstrong, David M. 1960. *Berkeley's Theory of Vision.* Parkville: Melbourne University Press.

Armstrong, David M. 1968. *A Materialist Theory of the Mind.* London: Routledge.

Atherton, Margaret. 1983. "Locke and the Issue Over Innateness." In *How Many Questions?,* edited by Isaac Levi, Charles Parsons, Leigh Cauman and Robert Schwartz, Indianapolis, Hackett. Reprinted in *Locke*, edited by Vere Chappell.

Berkeley, George. 1910. *A New Theory of Vision and Other Select Philosophical Writings.* Edited by Alexander Dunlop Lindsay. New York: Dutton.

Blackburn, Simon. 1990. "Filling in Space." *Analysis* 50, no. 2, 62-65.

Block, Ned. 1996. "Mental Paint and Mental Latex." *Philosophical Issues* 7, Perception, 19-49.

Block, Ned. 1997. "On a Confusion about a Function of Consciousness". In *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen Flanagan and Güven Güzeldere.

Block, Ned, Flanagan, Owen, and Güzeldere, Güven, eds. 1997. *The Nature of Consciousness: Philosophical Debates.* Cambridge, Mass.: MIT Press.

Block, Ned, and Robert Stalnaker. 1999. "Conceptual Analysis, Dualism, and the Explanatory Gap." *The Philosophical Review* 108, no. 1, 1-46.

Byrne, Alex. 2002. "Something About Mary." *Grazer Philosophische Studien* 63, 123-140.

Byrne, Alex. 2005. “Perception and Conceptual Content.” In *Contemporary Debates in Epistemology*, edited by Ernie Sosa and Matthias Steup. Malden: Blackwell.

Carr, David. 1979. “The Logic of Knowing How and Ability.” *Mind* 88, 351, 394-409.

Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Chalmers, David J. 2002. *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press.

Chalmers, David J. 2003. “The Content and Epistemology of Phenomenal Belief.” In *Consciousness: New Philosophical Perspectives*, edited by Quentin Smith and Aleksandar Jokic.

Chalmers, David J. 2004. “Phenomenal Concepts and the Knowledge Argument.” In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa and Daniel Stoljar.

Chalmers, David J., and Frank Jackson. 2001. "Conceptual Analysis and Explanation". *Philosophical Review* 110, 315-360.

Chappell, Vere, ed. 1994. *The Cambridge Companion to Locke*. New York: Cambridge University Press.

Chappell, Vere. 1994. “Locke's Theory of Ideas.” In *The Cambridge Companion to Locke*, edited by Vere Chappell. New York: Cambridge University Press.

Chappell, Vere, ed. 1998. *Locke*. New York: Oxford University Press.

Cheselden, William. 1727. “An Account of Some Observations Made by a Young Gentleman, Who Was Born Blind, or Lost His Sight so Early, That He Had no Remembrance of Ever Having Seen, and Was Couch’d between 13 and 14 Years of Age.” *Philosophical Transactions* 35, 447-450.

Churchland, Paul. 1985. "Reduction, Qualia, and the Direct Introspection of Brain States". *The Journal of Philosophy* 82, 8-28.

Churchland, Paul. 1989. “Knowing Qualia: A Reply to Jackson,” from his *A Neurocomputational Perspective.* Reprinted in *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa and Daniel Stoljar.

Conee, Earl. 1985. “Physicalism and Phenomenal Qualitities.” *The Philosophical Quarterly* 35, no. 140:296-302.

Cummins, Robert. 1978. “The Missing Shade of Blue.” *The Philosophical Review* 87, no. 4:548-565.

Dennett, Daniel C. 1988. “Quining Qualia.” Reprinted in *Mind and Cognition: A Reader*, edited by William Lycan.

Dennett, Daniel C. 1991. *Consciousness Explained.* Boston: Little, Brown and Co.

Descartes, Rene. 1988. *Descartes: Selected Philosophical Writings.* Translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch. New York: Cambridge University Press.

Dretske, Fred I. 2000. *Perception, Knowledge, and Belief: Selected Essays.* New York: Cambridge University Press.

Dretske, Fred I. 1999. *Knowledge and the Flow of Information.* Stanford, CA: CSLI Publications.

Evans, Gareth. 1982. *Varieties of Reference*. New York: Oxford University Press.

Fodor, Jerry A. 1998. *Concepts : Where Cognitive Science Went Wrong.* New York: Oxford University Press.

Fogelin, Robert J. 1984. “Hume and the Missing Shade of Blue.” *Philosophy and Phenomenological Research* 45, no. 2:263-271.

Gallagher, Shaun. 2005. *How the Body Shapes the Mind.* New York: Oxford University Press.

Garrett, Don. 1997. *Cognition and commitment in Hume's philosophy.* New York: Oxford University Press.

Gendler, Tamar Szabó, and Hawthorne, John, eds. 2006. *Perceptual Experience.* New York: Oxford University Press.

Gendler, Tamar Szabó, and Hawthorne, John, eds. 2002. *Conceivability and Possibility.* New York: Oxford University Press.

Ginet, Carl. 1975. *Knowledge, Perception, and Memory.* Dordrecht: Reidel.

Gunther, York H., ed. 2003. *Essays on Nonconceptual Content.* Cambridge, Mass.: MIT Press.

Hardin, C.L. 1988. *Colors for Philosophers.* Indianapolis: Hackett.

Harman, Gilbert. 1990. “The Intrinsic Quality of Experience.” *Philosophical Perspectives* 4, Action Theory and Philosophy of Mind, 31-52.

Heck, Jr. Richard J. 2000. “Nonconceptual Content and the ‘Space of Reasons’.” *The Philosophical Review* 109, no. 4:483-523.

Hellie, Benj. 2004. “Inexpressible Truths and the Allure of the Knowledge Argument.” In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa and Daniel Stoljar.

Horgan, Terence. 1984. “Jackson on Physical Information and Qualia.” *The Philosophical Quarterly* 34, no. 135:147-152. In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa and Daniel Stoljar.

Hume, David. 1739/2001. *A Treatise of Human Nature.* Edited by David Fate Norton and Mary J. Norton, New York: Oxford University Press.

Jackson, Frank. 1982. “Epiphenomenal Qualia.” *The Philosophical Quarterly* 32, no. 127:127-136. In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa and Daniel Stoljar.

Jackson, Frank. 1986. “What Mary Didn't Know.” *The Journal of Philosophy* 83, no. 5:291-295. In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa and Daniel Stoljar.

Jackson, Frank. 1998. “Postscript on Qualia.” In his *Mind, Methods, and Conditionals.* London: Routledge. Reprinted in *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa and Daniel Stoljar.

Jackson, Frank. 2003. “Mind and Illusion.” *Minds and Persons: Royal Institute of Philosophy Supplement:* 53. O'Hear, Anthony (ed). Reprinted in *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa and Daniel Stoljar.

Jolley, Nicholas. 1999. *Locke: His Philosophical Thought.* New York: Oxford University Press.

Kelly, Sean D. 2001a. “Demonstrative Concepts and Experience.” *The Philosophical Review* 110, no. 3:397-420.

Kelly, Sean D. 2001b. “The Non-Conceptual Content of Perceptual Experience: Situation Dependence and Fineness of Grain.” *Philosophy and Phenomenological Research* 62, no. 3:601-608.

Kind, Amy. 2003. “What's So Transparent about Transparency?” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 115, no. 3:225-244.

Koethe, John. 2002. “Stanley and Williamson on *Knowing How*.” Journal of Philosophy, vol. 99, no. 6, 325-328.

Leibniz, G. W. 1991. *Discourse on Metaphysics and Other Essays.* Indianapolis: Hackett.

Leibniz, G. W. 1765/1996. *New Essays on Human Understanding.* Edited by Peter Remnant and Jonathan Bennett, New York: Cambridge University Press.

Levin, Janet. 1986. "Could Love Be Like a Heatwave? Physicalism and the Subjective Character of Experience." *Philosophical Studies* 49, 245-261.

Levine, Joseph. 1983. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64, 354-361.

Levine, Joseph. 2001. *Purple Haze: The Puzzle of Consciousness.* New York: Oxford University Press.

Levine, Joseph. 2007. "Phenomenal Concepts and the Materialist Constraint." In *Phenomenal Concepts and Phenomenal Knowledge,* edited by Torin Alter and Sven Walters.

Lewis, David. 1988. "What Experience Teaches." In *Proceedings of Russellian* Society 13:29-57. Reprinted in *There's Something About Mary,* edited by Peter Ludlow, Yujin Nagasawa and Daniel Stoljar.

Loar, Brian. 1990/1997. "Phenomenal States." *Philosophical Perspectives* 4, no. Action Theory and Philosophy of Mind: 81-108. Revised and reprinted in *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen Flanagan and Güven Güzeldere.

Locke, John. 1691/1959. *An Essay Concerning Human Understanding.* Vol. Vol. 1. New York: Dover.

Ludlow, Peter, Nagasawa, Yujin, and Stoljar, Daniel, eds. 2004. *There's Something About Mary : Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument.* Cambridge, Mass.: MIT Press.

Lycan, William G., ed. 1990. *Mind and Cognition: A Reader*. Cambridge, Mass.: Blackwell.

Lycan, William G. 1996. *Consciousness and Experience.* Cambridge, Mass.: MIT Press.

Lycan, William G. 2003. “Perspectival Representation and the Knowledge Argument.” In *Consiousness: New Philosophical Perspectives*, edited by Quentin Smith and Aleksandar Jokic.

Margolis, Eric, and Laurence, Stephen, eds. 1999. *Concepts: Core Readings.* Cambridge, Mass.: MIT Press.

Martin, C. B., and Armstrong, D. M., eds. 1968. *Locke and Berkeley*. Notre Dame: University of Notre Dame Press.

McDowell, John H. 1994. *Mind and World.* Cambridge, Mass.: Harvard University Press.

Mellor, D. H. 1992/1993. “Nothing Like Experience.” *Proceedings of the Aristotelian Society* 93, 1-16.

Metzinger, Thomas, ed. 1995. *Conscious Experience.* Exeter: Imprint Academics.

Morgan, Michael J. 1977. *Molyneux's Question: Vision, Touch, and the Philosophy of Perception.* New York: Cambridge University Press.

Morreall, John. 1982. “Hume's Missing Shade of Blue.” *Philosophy and Phenomenological Research* 42, no. 3:407-415.

Nagel, Thomas. 1974. "What is It Like To Be a Bat?.” *Philosophical Review* 83, no. 4:435-450.

Nemirow, Laurence. 1980. “Book Review: Thomas Nagel’s *Mortal Questions*.” *The Philosophical Review* 89, no. 3:473-477.

Nemirow, Laurence. 1990. “Physicalism and the Cognitive Role of Acquaintance.” In *Mind and Cognition: A Reader,* edited by William Lycan.

Nida-Rümelin, Martine. 1995. “What Mary Couldn’t Know: Beliefs about Phenomenal States.” In *Conscious Experience*, edited by Thomas Metzinger.

Noë, Alva. 2004. *Action in Perception.* Cambridge, Mass.: MIT Press.

Noë, Alva. 2005. ***Analysis,*** Vol. 65 Issue 4, 278-290.

Parker, E.S., Cahill, L., McGaugh, G.L. 2006. “A case of unusual autobiographical remembering.” *Neurocase* 12, 1:35-49.

Peacocke, Christopher. 1992. *A Study of Concepts.* Cambridge, Mass.: MIT Press.

Peacocke, Christopher. 2001a. “Does Perception Have a Nonconceptual Content?” *The Journal of Philosophy* 98, no. 5:239-264.

Peacocke, Christopher. 2001b. “Phenomenology and Nonconceptual Content.” *Philosophy and Phenomenological Research* 62, no. 3:609-615.

Perry, John. 2001. *Knowledge, Possibility, and Consciousness*. Cambridge: MIT Press.

Pitt, David. 2004. “The Phenomenology of Cognition, or, What is it Like to Think That P?” *Philosophy and Phenomenological Research*, LXIX, 1-36.

Prinz, Jesse. 2002. *Furnishing the Mind Concepts and their Perceptual Basis.* Cambridge: MIT Press.

Raffman, Diana. 1993. *Language, Music, and Mind.* Cambridge, Mass.: MIT Press.

Raffman, Diana. 1995. “On the Persistence of Phenomenology” in *Conscious Experience*, edited by Thomas Metzinger, 293-308.

Riskin, Jessica. 2002. *Science in the Age of Sensibility: The Sentimental Empiricists of the French Enlightenment.* Chicago: University of Chicago Press.

Rosenthal, David M. 1986. “Two Concepts of Consciousness.” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 49, 329-359.

Rosenthal, David. 1997. “A Theory of Consciousness.” In *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen Flanagan, and Güven Güzeldere, 729-753.

Rumfitt, Ian. 2003. “Comments and Criticism: Savoir Faire.” Journal of Philosophy, vol. 100, no. 3, 158-166.

Russow, Lilly-Marlene. 1980. “Simple Ideas and Resemblance.” *The Philosophical Quarterly* 30, no. 121:342-350.

Ryle, Gilbert. 1933. “John Locke on the Human Understanding.” In *Locke and Berkeley*, edited by C. B. Martin and D. M. Armstrong.

Ryle, Gilbert. 1949. *The Concept of Mind.* London: Hutchinson.

Schwitzgebel, Eric. 2002. "How Well Do We Know Our Own Conscious Experience? The Case of Visual Imagery". *Journal of Consciousness Studies* 9, no. 5-6:35-53.

Schwitzgebel, Eric. 2006. "Do Things Look Flat?” *Philosophy and Phenomenological Research* 72, 589-599.

Sepper, Dennis L. 1996. *Descartes's Imagination*. Berkeley: University of California Press.

Shoemaker, Sydney. 1984. “Churchland on Reduction, Qualia, and Introspection.” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1984, no. Volume Two: Symposia and Invited Papers:799-809.

Shoemaker, Sydney. 1994a. “Phenomenal Character.” *Nous* 28, no. 1:21-38.

Shoemaker, Sydney. 1994b. “Self-Knowledge and ‘Inner Sense’ Lecture I: The Object Perception Model.” *Philosophy and Phenomenological Research* 54, no. 2:249-269.

Shoemaker, Sydney. 1994c. “Self-Knowledge and ‘Inner Sense’: Lecture II: The Broad Perceptual Model.” *Philosophy and Phenomenological Research* 54, no. 2:271-290.

Shoemaker, Sydney. 1994d. “Self-Knowledge and ‘Inner Sense’: Lecture III: The Phenomenal Character of Experience.” *Philosophy and Phenomenological Research* 54, no. 2:291-314.

Smith, Quentin, and Jokic, Aleksandar, eds. 2003. *Consciousness: New Philosophical Perspectives.* New York: Oxford University Press.

Snowdon, Paul F. 2004. “Knowing How and Knowing That: A Distinction Reconsidered.” *Proceedings of the Aristotelian Society* 104, 1-29.

Speaks, Jeff. 2005. “Is There a Problem about Nonconceptual Content?” *Philosophical Review* 114, no. 3:359-398.

Spelke, Elizabeth S. 1998. “Nativism, Empiricism, and the Origins of Knowledge.” *Infant Behavior and Development* 21, no. 2:181-200.

Stalnaker, Robert. 1998. “What Might Nonconceptual Content Be?” *Philosophical Issues* 9, no. Concepts:339-352.

Stanley, Jason, and Williamson, Timothy. 2001. *Journal of Philosophy*. 98, 8, 411-444.

Stoljar, Daniel. 2000. “Physicalism and the Necessary a Posteriori.” *The Journal of Philosophy* 97, no. 1:33-54.

Stoljar, Daniel. 2001. “Two Conceptions of the Physical.” Reprinted in abbreviated form in *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa and Daniel Stoljar.

Stoljar, Daniel. 2005. “Physicalism and Phenomenal Concepts.” *Mind and Language* 20, no. 5:469-494.

Thau, Michael. 2002. *Consciousness and Cognition*. New York: Oxford University Press.

Tye, Michael. 1991. *The Imagery Debate.* Vol. xiv. Cambridge, Mass.: MIT Press.

Tye, Michael. 2002. *Consciousness, Color, and Content.* Cambridge, Mass.: MIT Press.

Tye, Michael. 2006. “Nonconceptual Content, Richness, and Fineness of Grain.” In *Perceptual Experience* edited by Tamar Szabó Gendler and John Hawthorne.

Van Gulick, Robert. 2004. “So Many Ways of Saying No to Mary.” In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa and Daniel Stoljar.